

GenCore version 5.1.4.P5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:39:55 ; Search time 32.9328 seconds

(without alignments)
3666.364 Million cell updates/sec

File: US-10-092-390-4

Perfect score: 3601
Sequence: 1 MVISLNSCLSFICLLCHWLT.....HCDSYCAEGRMGPNCLPCY 586

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SPREMBL_21:
1: sp_archaea:
2: sp_bacteria:
3: sp_fungi:
4: sp_human:
5: sp_invertebrate:
6: sp_mammal:
7: sp_mhc:
8: sp_organelle:
9: sp_plant:
10: sp_protist:
11: sp_virus:
12: sp_vertebrate:
13: sp_unclassified:
14: sp_virus:
15: sp_bacteriophage:
16: sp_bacteriophage:
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	3601	100.0	1140 4 096K7	096K7 homo sapien
2	3468	96.3	567 4 08WU3	08WU3 homo sapien
3	2094	58.2	969 4 096K6	096K6 homo sapien
4	1828	50.8	747 11 08VH4	08VH4 mus musculu
5	1824	50.7	1034 11 08VH7	08VH7 mus musculu
6	1824	50.7	1034 11 08VH5	08VH5 mus musculu
7	1372.5	38.1	1574 11 088281	088281 ratu mus
8	1340	37.2	1664 5 09VQ2	09VQ2 caenorhabdi
9	1309	36.4	220 11 063404	063404 ratu mus
10	1284.5	35.7	1045 5 08T36	08T36 caenorhabdi
11	1284.5	35.7	1070 5 08T37	08T37 caenorhabdi
12	1284.5	35.7	1111 5 09XW6	09XW6 caenorhabdi
13	1282.5	35.6	1246 4 075095	075095 homo sapien
14	1252	34.8	434 5 09W0A0	09W0A0 drosophila
15	741	20.6	594 5 09W0A1	09W0A1 drosophila
16	740	20.5	594 5 09Y151	09Y151 drosophila

17	708.5	19.7	2447	13 013149	013149 fuqu rubrip
18	685.5	19.0	2524	5 09GPA5	09GPA5 brachiolesto
19	682	18.9	4288	4 09NFK9	09NFK9 homo sapien
20	681.5	18.9	2531	5 016004	016004 lytechinus
21	681	18.9	2653	5 025253	025253 lucilia cup
22	678	18.8	1193	13 090819	090819 gallus gall
23	676.5	18.8	4006	11 035452	035452 mus musculu
24	672.5	18.7	4135	6 018977	018977 bos taurus
25	667.5	18.5	4114	11 054796	054796 mus musculu
26	664	18.4	1214	13 0907D2	0907D2 xenopus lae
27	648.5	18.0	1254	13 09YH02	09YH02 brachydanio
28	648.5	18.0	1254	13 090Y56	090Y56 brachydanio
29	648.5	18.0	2352	5 061240	061240 haemaphysa
30	646.5	18.0	713	5 096289	096289 podocoryne
31	646	17.9	1532	13 090984	090984 gallus gall
32	646	17.9	1714	13 090995	090995 gallus gall
33	641.5	17.8	2019	11 064706	064706 mus musculu
34	638	17.7	3695	4 08TDF8	08TDF8 homo sapien
35	636	17.7	1810	13 090824	090824 gallus gall
36	628.5	17.5	1216	13 090Y55	090Y55 brachydanio
37	622.5	17.3	2146	5 09VC97	09VC97 drosophila
38	621	17.2	752	13 042374	042374 brachydanio
39	609	16.9	3319	5 09VJ75	09VJ75 drosophila
40	609	16.9	3367	5 09XZC9	09XZC9 drosophila
41	595	16.5	1212	13 042347	042347 gallus gall
42	593.5	16.5	3704	5 091904	091904 caenorhabdi
43	586	16.3	1722	5 019350	019350 caenorhabdi
44	583.5	16.2	2656	5 09GN03	09GN03 paracentrot
45	581	16.1	1404	5 09VB65	09VB65 drosophila

ALIGNMENTS

RESULT 1	ID	096K7	PRELIMINARY	PRT	1140 AA.
AC	096K7				
DT	01-DEC-2001 (TREMBLrel. 19, Created)				
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	MEGF10 protein (K1A1780).				
GN	MEGF10.				
OS	Homo sapiens (Human).				
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=HIPPOCAMPUS;				
RX	MEDLINE=24245430; PubMed=11347906;				
FA	Nagase T., Nakayama M., Nakajima D., Kikuno K., Ohara O.;				
RT	The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.				
RT	GENBANK=U00000.1				
DR	EMBL; AB058676; BABF7409.1;				
DR	InterPro; IPR000561; EGF-like.				
DR	Pfam; PF00008; EGF_15.				
DR	PROSITE; PS00022; EGF_1; UNKNOWN_17.				
DR	PROSITE; PS01186; EGF_2; UNKNOWN_17.				
SO	SEQUENCE 1140 AA; 122204 MW; 4552FA239423895A CRC64;				
Query Match	100.0%; Score 3601; DB 4; Length 1140;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
OY	1	MVISLNSCLSFICLLCHWLTASPLNEDNVCSHMSVSYVQESYPRHFDIYITSC	60		
Db	1	MVISLNSCLSFICLLCHWLTASPLNEDNVCSHMSVSYVQESYPRHFDIYITSC	60		
OY	61	TDIINWFCTHRVSYRATVYRHEGKTVRRSQCPCPGYEGKCVPCADKCVHGACIA	120		

```

Db 61 TDILMFKCTRRHVSRYTRAYRHGEKTMRYRKQCCPGFESEEMCVPHCADXCVRHCRTA 120
Oy 121 PNTCOCEPFGMGSTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCED 180
Db 121 PNTCOCEPFGMGSTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCED 180
Oy 181 RCEOGTYGNDCHQRCOCQNGATCDHYTGBECRCRPGYTGAFCEDLCPGKHGPOCEQRCP 240
Db 181 RCEOGTYGNDCHQRCOCQNGATCDHYTGBECRCRPGYTGAFCEDLCPGKHGPOCEQRCP 240
Oy 241 QNGCVCHNHTGBCSCPSGSMGTVCQPCPEGRGKNCSECCCHNGTCDATGQCHCSP 300
Db 241 QNGCVCHNHTGBCSCPSGSMGTVCQPCPEGRGKNCSECCCHNGTCDATGQCHCSP 300
Oy 301 GTTGERCODECVGTGYVCAETCCQVNGKCYHVSAGALCEAGFAGRCERARLCEGLY 360
Db 301 GTTGERCODECVGTGYVCAETCCQVNGKCYHVSAGALCEAGFAGRCERARLCEGLY 360
Oy 361 GIKCDKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETCSPGFYGEACQOICSCQNGAD 420
Db 361 GIKCDKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETCSPGFYGEACQOICSCQNGAD 420
Oy 421 CDSYTKCTCAGPFGKIDCSTPCPLGTGYGINSRCCKNDVAVSPYDSCCTCKAGMHGV 480
Db 421 CDSYTKCTCAGPFGKIDCSTPCPLGTGYGINSRCCKNDVAVSPYDSCCTCKAGMHGV 480
Oy 481 DCSIRPSTGTFWFGCNLTCCCLNGGACNTLDGTCTCAPMRBEKELPCODGTGYNCAE 540
Db 481 DCSIRPSTGTFWFGCNLTCCCLNGGACNTLDGTCTCAPMRBEKELPCODGTGYNCAE 540
Oy 541 RCDCHADCGHPTTGHCRLPGHSGYHCDGVCAEGRWGNCISLPCY 566
Db 541 RCDCHADCGHPTTGHCRLPGHSGYHCDGVCAEGRWGNCISLPCY 566

RESULT 2
Oy 08WUL3 PRELIMINARY: PRT: 567 AA.
AC 08WUL3;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Similar to MEGF10 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC020198; AAH20198.1; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS; PR00011; EGF_LAMININ.
DR SMART; SM00181; EGF_11.
DR SMART; SM00180; EGF_Lam; 9.
DR PROSITE; PS00022; EGF_1; UNKNOWN_10.
DR PROSITE; PS01186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 567 AA; 60797 MW; CTFBBDCEB7C627 CRC64;

Query Match 96.3%; Score 3468; DB 4; Length 567;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 565; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 MYISLNSCSFTCLLCHITGASPLNEDPVNCSHMEISYTVQESYHPEDQIYITSC 60
Db 1 MYISLNSCSFTCLLCHITGASPLNEDPVNCSHMEISYTVQESYHPEDQIYITSC 60
Oy 61 TDLMMFKCTRRHVSRYTRAYRHGEKTMRYRKQCCPGFESEEMCVPHCADXCVRHCRTA 120
Db 61 TDLMMFKCTRRHVSRYTRAYRHGEKTMRYRKQCCPGFESEEMCVPHCADXCVRHCRTA 120

```

```

Db 61 TDILMFKCTRRHVSRYTRAYRHGEKTMRYRKQCCPGFESEEMCVPHCADXCVRHCRTA 120
Oy 121 PNTCOCEPFGMGSTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCED 180
Db 121 PNTCOCEPFGMGSTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCED 180
Oy 181 RCEOGTYGNDCHQRCOCQNGATCDHYTGBECRCRPGYTGAFCEDLCPGKHGPOCEQRCP 240
Db 181 RCEOGTYGNDCHQRCOCQNGATCDHYTGBECRCRPGYTGAFCEDLCPGKHGPOCEQRCP 240
Oy 241 QNGCVCHNHTGBCSCPSGSMGTVCQPCPEGRGKNCSECCCHNGTCDATGQCHCSP 300
Db 241 QNGCVCHNHTGBCSCPSGSMGTVCQPCPEGRGKNCSECCCHNGTCDATGQCHCSP 300
Oy 301 GTTGERCODECVGTGYVCAETCCQVNGKCYHVSAGALCEAGFAGRCERARLCEGLY 360
Db 301 GTTGERCODECVGTGYVCAETCCQVNGKCYHVSAGALCEAGFAGRCERARLCEGLY 360
Oy 361 GIKCDKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETCSPGFYGEACQOICSCQNGAD 420
Db 361 GIKCDKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETCSPGFYGEACQOICSCQNGAD 420
Oy 421 CDSYTKCTCAGPFGKIDCSTPCPLGTGYGINSRCCKNDVAVSPYDSCCTCKAGMHGV 480
Db 421 CDSYTKCTCAGPFGKIDCSTPCPLGTGYGINSRCCKNDVAVSPYDSCCTCKAGMHGV 480
Oy 481 DCSIRPSTGTFWFGCNLTCCCLNGGACNTLDGTCTCAPMRBEKELPCODGTGYNCAE 540
Db 481 DCSIRPSTGTFWFGCNLTCCCLNGGACNTLDGTCTCAPMRBEKELPCODGTGYNCAE 540
Oy 541 RCDCHADCGHPTTGHCRLPGHSGY 566
Db 541 RCDCHADCGHPTTGHCRLPGMSG 566

RESULT 3
Oy 096KG6 PRELIMINARY: PRT: 969 AA.
AC 096KG6;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE The complete sequences of 100 new cDNA clones from brain which code
DE for large proteins in vitro.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE-21245130; PubMed-11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL EMBL; AB058677; BAB47410.1; -
RL DNA Res. 8:85-95(2001).
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001281; Rleske.
DR Pfam: PF00008; EGF_15.
DR PROSITE; PS00022; EGF_1; UNKNOWN_17.
DR PROSITE; PS01186; EGF_2; UNKNOWN_17.
DR PROSITE; PS00200; RIESKE_2; UNKNOWN_1.
SQ SEQUENCE 969 AA; 101600 MW; 56DDJFEEL39C8209 CRC64;

Query Match 58.2%; Score 2094; DB 4; Length 969;
Best Local Similarity 65.4%; Pred. No. 2,2e-194;
Matches 312; Conservative 58; Mismatches 107; Indels 0; Gaps 0;

```

```

Oy 109 CADKVGRCIAPNTCCPFGNGTNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACH 168
Db 28 CTECVHGRVSDPTDCHCEPFGMGDPDSCSDHMGPHCSNRCCCKNGALCNPTTGACH 87

```


Db	65	QPTVYRRTVYQVVKKMDSRRLQDCRRGTYEERAGCVFLCQNECHGNCVAPNQCQCARGW	124
OY	131	GGTNCSSACDGDHWGPHCTSHCKOCNGALCNPTTGACHCAAGFRGWRCEDRCEDEGTGND	190
Db	125	RGGCCSSCACAGMMGPOCDKRCFHCGNNSSCDPCSGACFCPSGLDPPNCLAPCRAGHTGPA	184
OY	191	CHQRCOCQNGATCDHYVGECCRCRGTYGARCEDELCPPRGKHGRPOEJQRCPOQNGVCHNVT	256
Db	185	CFQDQCY-GASCPQDQACCPGRGARGPSCNVPCSGTQDGFEPRIYPCQNGGVVPGSQ	243
OY	251	GECSCPGSMNMTYVCQGPCPEGRFPRKNSOECCCHNGGTCAATGQCSCPSGYTERCODE	310
Db	244	GSCGCPPEMMKVIISLPREBGFHNPCTOEBCRHNGGLCDRFYQGHCAAGYIDRKQEE	303
OY	311	CPVETGYVLCAETQCVCVNGGFCYHVSQACLCCEAGFAGERCEARLCPEGLYIKDKRKPC	370
Db	304	CPVRFQGDCCAEETDCAPGARCFPANAACLCIEHGFETDRCETRLCPRGGRGLSCQEPCTC	363
OY	371	HLEMTSHPRMSSGACAPKPGSGSLXCMETSPPRGTYEERACQOICSCQNGADODSYTKCTC	430
Db	364	DPEHLSCHPRHNGECSQCPGNAAGLHNCBSCQDTHNGRCEHCILTLHGGLCAGSLCRC	423
OY	431	APRFGKIDCSTPRCELTGYTGINCSSRCCCKNDVAVSPYDQSCCTCAGAHGYDCSIRPCSGT	490
Db	424	APGYTGPRCANLCPRPDTYTGINCSSRCCSENNINASPIDTGTICKEGMRGNCISVPCPLGT	483
OY	491	WFGCNLTCCCLNGAGACNTLDGTCTCARGWRGEKCELLPCQDGTGYGLNCAERCDCHADGC	550
Db	484	WGFNMNASCOCAGHGVCSFQTAGACTCTPGMHGAJCOLPCRGQFGECCASVCDCHSDGC	543
OY	551	HPTGHCKCLPAGMSGVHDCSACAGRRGNPNCSLPC	585
Db	544	DVPHGQCRQAGMMGTGRCHLPCPRGTFMGANCSNTC	578

```

RESULT 6
08VTKS ID Q8VIK5 PRELIMINARY; PRT; 1034 AA.
AC Q8VIK5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MEGF12.
GN MEGF12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclauognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Ivanova N.B., Lemischka I.R.;
RT "The global gene expression profiling of the hematopoietic stem
    cell."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440279; AAL3583.1; .
DR InterPro; IPRO00561; EGF-like.
DR Interpro; IPRO02049; LamInln_EGF.
DR Pfam; PF00008; EGF_9.
DR PRINTS; PR00011; EGFLAMININ.
DR SMART; SM00181; EGF; 15.
DR SMART; SM00180; EGF_Lam; 14.
DR PROSITE; PS00022; EGF_1; UNKNOWN_13.
DR PROSITE; PS01186; EGF_2; UNKNOWN_12.
SO SEQUENCE 1034 AA; 110580 MB; 714E5016848BE4C CRC64;

Query Match          50.7%; Score 1824; DB 11; Length 1034;
Best Local Similarity 51.8%; Pred. No.3,4e+168;
Matches 298; Conservative 59; Mismatches 212; Indels 6; Gaps 3;

```

Db	7	LLALGLRLGTLSNDPNCVCTFMESEPTTYYTKSHLRPESLPAESCH--RPMEDPHTCA	64
Qy	71	RHRVSYTRATPRHGEKIMYRRKSQCCPEFYSGSECVPRHCADKCVHGRCTAPNTCCSEPM	130
Db	65	QPTVYRTYVYKQVKMDSRRPLCCSCRGYSESRGCAVPLCAOEVBHGRCAVPANCCOCAPGM	124
Qy	131	GGTNCSSACOSDHHGPHCTSRCCCKNCLNPLTGACHCAAGFRGRCBDBEOGYGMD	190
Db	125	RGGGSSSECAEGMMGPCCDKFCHIGNNNSCDPRASGTCFCSGSLDPPNCLDPCAHGYPA	184
Qy	191	CHORCCONGATCDHYTGECRCRPGYVGAECEDLCPRGKHGPOCEORCCQNGVCHNT	250
Db	185	CQPDQCY-CASCDPRQGAFCFRPGRGRPCSNTPCSQGTGDFEFCRPRTYCCQNGVYQSSQ	243
Qy	251	GECSCPGSMGTVGQDRCPEGRFEGKNCSDQCHNGSTDAATGQCHSPGYTGERCODE	310
Db	244	GSQSCPRGMWGVISLPCRPFGFHNPCTQBCRHNGGLCDRFTGQCHCAPYIGDRCEE	303
Qy	311	CPVGTGYVLCAETQCVCYNGGKCHVSNACLCBAAGFAGEREALCEBGLYGIKDKRCPC	370
Db	304	CPVRFQGDCAETDCAKPGARCFRPAFACALCEHGFGRDCTELCDGRGSLSCDEPTC	363
Qy	371	HLEMYCHPMNSGECAKCPKPSGLYNETCSPEFYGBACAOICSCNGADCSYTSKTC	430
Db	364	DPEHSLSCHPMHGECSQCPQFNAGLHCHESCPRODTHGBCQENHCLHGLCLADSLCLPC	423
Qy	431	APGFKGIDCSYPCPLGTGYGINCSSRCCKNDAYACSPVDSCTCKAGMHGVDCSIRCPST	490
Db	424	APGYTGPICAMLCPRDYTGIGINCSSRCSENAIACSPIDGTCLCKBEQMOGNSVCPRLT	483
Qy	491	WFGQCLTLTQCLNGACNTLDGTCTCAPGRGKEKELPCODGYGLNCAERCDCSHADGC	550
Db	484	WGFNMCNACQCAHGVCSPTGTACTCTPGMNGHACQLPCPKGQFGEGCASVCDCHSDGC	543
Qy	551	HTTGHCRCLEPGMSGVHCDVSACAEGRGRPCSLPC	585
Db	544	DVHGOQRCQAGMWGTRCHLPCPEGEFGANCSNTC	578

RESULT	7
ID	088281
PRELIMINARY;	PRT; 1574 AA.
AC	088281;
DT	01-NOV-1998 (TREMBLrel. 08, Created)
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	MEGF6.
GN	MEGF6.
OS	Rattus norvegicus (Rat.).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBI_TaxID=10116;
RA	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX	MEDLINE=98360089; PubMed=9693030;
RA	Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Chara O.;
RT	"Identification of high-molecular-weight proteins with multiple EGF-
RT	like motifs by motif-trap screening."
RL	Genomics 51:127-34(1998).
RL	EMBL; AB011532; BAA32462.1; .
DR	HSSP; P00736; IAPQ.
DR	InterPro: IPR000152; Asx_hydroxyl.
DR	InterPro: IPR000561; EGF-like.
DR	InterPro: IPR001881; EGF_Ca.
DR	Pfam; PF00008; EGF_24.
DR	SMART; SM00179; EGF_Ca; 4, 19.
DR	SMART; SM00001; EGF_like; 19.
DR	PROSITE; PS00010; ASX_HYDROXYL; 5.
DR	PROSITE; PS00022; EGF_1; UNKNON_23.
DR	PROSITE; PS0186; EGF_2; 23.
DR	PROSITE; PS0187; EGF_Ca; 5.
FW	Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.

SO SEQUENCE 1574 AA; 165445 MW; 28485338B8F77E67 CRC64;

Query Match 38.1%; Score 1372.5; DB 11; Length 1574;

Best Local Similarity 41.3%; Pred. No. 2.8e-124; Matches 229; Conservative 56; Mismatches 198; Indels 71; Gaps 9;

QY 94 CCEPGESEGCYVHCADKCVHG-----RCIAPN-----TCCEPGMGNTNCS 136
 DB 808 CLPFGVS-----RCDDTSAGWYGTGCOIRACAGNDGHDPTTGKSCAPRGMTGLSCQ 861
 QY 137 SAGDGHMGPHTSRTCCCKNG-ALCNDPTGACACAGFRGWRCEDEQGTGGNDCHQRC 195
 DB 862 RACDSGHMGPDCIHPNCNSAGHGNCDAVSGICLCEAGYEGRCBQSCROGGYGGSCBQKC 921
 QY 196 QCCGATCDHYTGRCRPPGTGAFCD-----LC 225
 DB 922 RCEGAACDHYSGACTCPAGWRGSCFHACPAGEFGILDSCSACNSAGAPCDAYTGSIC 981
 QY 226 PGRKHGQCEBQRC-----CONGVCHHYTGECSCPGMWGTYCGQPCPEGR 272
 DB 982 PAGRWGRCAQSCPLTFGLNCSQICTFNASCDSSTGQCHCAPGMGPTCLDACPGL 1041
 QY 273 FGKNSQSECOHNGTCDAAATGQCHCSPGYTGECRQDECPVGTGVLCATCQCVNGKC 332
 DB 1042 YGKNCQHSCLCRNGRCDDPILGQCTCEGWTGLACNECLPGHYAAGCQLNCSCLHGIC 1101
 QY 333 YHVGALCLERAGFERGCEARLCEGLXGKDKRCRCHLENTHSCHPMSECCAKRGWS 392
 DB 1102 DRLGHCLCPAGWTDGDCSS-CVSGTFGVHCEHCAC--RKGASCHHYTGACPCPGWR 1158
 QY 393 GLYCNETSPGFYEACQOICSCONGADCSVTGKCTCAGCFKIDSTPCPLGTGIC 452
 DB 1159 GPRHQAQPRGWRGCAQRCICPTNASHHYTGECRCRPGFTGLSEBQACOPTEFKDC 1218
 QY 453 SSRGCKNDA-VCSVDGSCCTCAAGHGVDCSIRCPSTGWFPGCNLTQCLNGSACNTLD 511
 DB 1219 EHLQCCGEGEWACDPASGVCTCAAGYHGTGCLQRCPSGRGRCGHEHCKICNGSTCDPAT 1278
 QY 512 GTCTCAGWGEKELPCODGTGVLNCAERDCSHADGCHPTTGCHCCLTGMGSHVHDSV 571
 DB 1279 GACTCPRGFLGADSLCPGGRFPSCAHYACRQGAACDPSGACICSPKGTGVRCEHG 1338
 QY 572 CAEGRMGPNCLPC 585
 DB 1339 CPQDRFGKGCCLKC 1352

RESULT 8
 ID 09TV02 PRELIMINARY; PRT; 1664 AA.
 AC 09TV02;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Y64G10A.7 protein.
 GN Y64G10A.7.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderiinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN SEQUENCE FROM N.A.
 RA Mortimore B.J.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 RP [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RL Investigating biology".
 RL Science 282:2012-2018(1998).
 RP [3]
 RP SEQUENCE FROM N.A.

RA Ainscough R.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL117206; CAB60454.1; JOINED.
 DR EMBL; AL110498; CAB60454.1; JOINED.
 DR EMBL; AL110498; CAB57911.1; JOINED.
 DR EMBL; AL117206; CAB57911.1; JOINED.
 DR HSSP; P00736; IAPQ.
 DR InterPro; IPR00152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF_25.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00179; EGF_CA; 4.
 DR SMART; SM00001; EGF_Like; 18.
 DR PROSITE; PS00010; ASX_HYDROXYL; 4.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_22.
 DR PROSITE; PS01186; EGF_2; 24.
 DR PROSITE; PS01187; EGF_CA; 3.
 KW Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
 SO SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 37.2%; Score 1340; DB 5; Length 1664;
 Best Local Similarity 36.4%; Pred. No. 4.2e-121;
 Matches 237; Conservative 70; Mismatches 218; Indels 126; Gaps 13;

QY 16 LCHRI-----GTASPL-----NLEDPNCSHNESTVYQESYRHPFDIYITSC 60
 DB 824 VCHVTGTCTCLPQKTPCLDQSCAPNTYGPN-CAH-----TC 860
 QY 61 TDILMFCTRHRYSTAYRHEKTYRRKRSQCCPGEYSEGCYVHCAD-----TC 111
 DB 861 S-CYNGAKCESDS-----CHCTPFTY--GATCSEVCTPGRGIDCMQ 901
 QY 112 --KCVHR-GIAPN-TCQCEPGWGTNCSACDGDHMGPHCTSRQCKNALCNPTTGAC 167
 DB 902 LCKQNGAICDTSNGSECAPAGMGKRCDAAPGTGKDCSKKCDADMHCDPSPGEC 961
 QY 168 HCAAGFGMGCEDEGEGTGNDCHQCCQNGAT----- 202
 DB 962 ICPGKRGKHCDCTGSDGLFAGCKGICSCQNGATCDSTVSGCECRPGMRGKKCDRCPD 1021
 QY 203 -----CDHYTGECRCPPGYTGAFCECLCPGKHGPOC 234
 DB 1022 GRREGCNALCDDCTTNDTSMYNPFAKCDHYTGECRCPPGYTGAFCECLCPGKHGPOC 1081
 QY 235 EQRCPQNGGVCHHYTGECSCPSGMGTVCQGPCPEGRFGKNSQECQCHNGGTCDATG 294
 DB 1082 RHSCQCSNGASCDRVTGFCDCPSGFMKNCSECEPGLMGSNCMKHCLMHHGECNKENG 1141
 QY 295 QCHCSPTYTEKCODECPVGTGVLCATCQCVNAGGCIHYSGACLCFAFAGRCARL 354
 DB 1142 DCEIDIDMTSPSEFLCPFOFGFNCAORCNKNGASCADRKTGCECLPMSGHEHC-KS 1200
 QY 355 CPEGLYGIKCDKRCPCHEWTHSCHPMSEGCACRPGSGLYGNTCSPGFYGEACQOICS 414
 DB 1201 CVSHYAKCEBTECE--ENGALCDPISGHCSCQPGKRGKRCPLAKGFRGHCSSCR 1258
 QY 415 CQNGADCSVTGKCTCAGFGKIDGSTPCPLGTGVLNCSRGCGCKNDVAVSPVDGSCYCK 474
 DB 1259 CANSKSCDHTISGRQCPKGYAGHSCTELCPDFTGEGSCSQKCCDGGESMDAISGKFCFK 1318
 QY 475 AGHGVDCSIRCPSTGTFEGCNLTQCLNGACNTLDGTCTCAGRWGEKELPCQDGT 534
 DB 1319 PGHSGSDCKSGGVQGRGPPCNDLSCENGVCDDSTGSCVCPGPGYGTGTCELAQSDRF 1378
 QY 535 GLNCAERDCSHADGCHPTTGCHRCPLGMSGVHCDSDVACAGRMGPNCLPC 585
 DB 1379 GPFCERKCNENGTCDRLTGQRCPLPFTGMCNQCVPBGRGAGCKENC 1429

RESULT 9
 063404

ID	AC	PRELIMINARY;	PRT;	220 AA.
063404:				
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)		
OS	(Clone REM4) OR (Fragment).			
DE	Rattus norvegicus (Rat).			
OC	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HOLZMAN; TISSUE=BRAIN;			
RX	MEDLINE=9623515; Pubmed=864059;			
RA	Aaskura K., Poggali R.J., Pease L.R., Rodriguez M.;			
RT	remyelination is highly polyreactive to multiple known and novel			
RT	antigens.;			
RL	J. Neuroimmunol. 65:11-19(1996).			
RR	EMBL; L4186; BAB05844.1; .			
RR	HSP; P01132; IEGF.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR002049; Laminin_EGF.			
DR	Pfam: PF00008; EGF_3.			
DR	PRINTS; PR00011; EGF_LAMININ.			
DR	SMART; SM00180; EGF_Lam; 2.			
DR	SMART; PS00022; EGF_1; UNKNOWN_5.			
DR	PROSITE; PS0186; EGF_2; 5.			
KW	EGF-like domain; Glycoprotein.			
FT	NON_TER			
FT	NON_TER			
SS	SEQUENCE	220 AA; 23231 MW; 3119D391EAF64372 CRC64;		

Query Match	36.4%;	Score 1309;	DB 11;	Length 220;
Best Local Similarity	95.4%;	Pred. No. 5.8e-119;		
Matches 209; Conservative	4;	Mismatches 6;	Indels 0;	Gaps 0

QY	200	GATDHTATGECRCRPVTGAAFCEDLCSPRKNGPCCEDRCRCSCNGVCHNHTGTSCSPSG	253
Db	2	GATDHTATGECRCRPVTGAAFCEDLCSPRKNGPCCEDRCRCSCNGVCHNHTGTSCSPSG	61
QY	260	MGTVCGPCPEGRGKNCSDCCCHNGGTDAATGQCCHSPYNGERCDCPCPGTGYVL	319
Db	62	MGTVCGPCPEGRGKNCSDCCCHNGGTDAATGQCCHSPYNGERCDCPCPGTGYVL	121
QY	320	CAETCCQVNGGCKYHVSAGCLCEAGFAGERCEARLCBGLYGIKCDKRCPCHEINTHSCH	379
Db	122	CAETCCQVNGGCKYHVSAGCLCEAGFAGERCEARLCBGLYGIKCDKRCPCHEINTHSCH	181
QY	380	PMSEGCACKPQMSGLYCNENCSPPFYEDACQDITSCNG	418
Db	182	PMSEGCACKPQMSGLYCNENCSPPFYEDACQDITSCNG	220

RESULT	10			
08RT3A6				
ID	08RT3A6	PRELIMINARY:	PRT:	1045 AA.
AC	08RT3A6;			
DT	01-JUN-2002 (TrEMBLrel. 21, Created)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Y47H9C.4c protein.			
GN	Y47H9C.4C.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae.			
OC	Rhabditidae; Palodermineae; Caenorhabditis.			
OX	NCBI_TaxID=62339;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Harris B.R.;			
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.			
RL	[2]			

RP SEQUENCE FROM N. A.
RX MEDLINE-99069613; PubMed-9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology.";
RL Science 283:2012-2018(1998).
DR EMBL: AL032657; GDB27615.1; -.
SQ SEQUENCE 1045 AA; 111723 MW; 0A6D57A3A80BCDCA CR664;

Query Match	35.7%;	Score 1284.5;	DB 5;	Length 1045;
Best Local Similarity	34.2%;	Pred. No. 6.3e-116;		
Matches 246;	Conservative 77;	Mismatches 221;	Indels 175;	Gaps 20

OY	21	GTASPLNEDNEDYNSHMSYSVATVOESYPPHFPDDIYT-----SCTDILNNEKCYRHR	73
Db	35	GTEP-----QGBHVC-----VKTIYDDX-----ELKVIHTVYVNDTEOCLNPLTJGPOC----	80
OY	74	VSRYRATYHNGKTYRRK-----SQCCEFESESE-NCVPHCADKCYHNGRIANTC	1244
Db	81	----TVERKGOASTYOROLVAKERYKVCSCGGYYOTDHNHCLPCPNPPCKKGIETEPKC	1368
OY	125	OCEPMGSGTNCSSACDDGDMHPHCTSRQCKNGALNPIYAGACHAAGFGRMRE-----	1797
Db	137	EODPEYGGKTYCASSCSVGTWTLGSKSODCGNANCPDELGTCTCTSGYFGEREKPCPD	1966
OY	180	-----DCEBQTYGMDCHOROCONGAT	2022
Db	197	NKMGNVCYKSCPCONGCKNEKRCVCSDDGGEFCLNKCEEGKFTGAECKFEFCMONGAT	2556
OY	203	CDHYMGEBCRCPRGYTGACEDLCPGKHKPGCEBRCPCONGCVNHYHTGCSOSGMMGT	2652
Db	257	CDNTNGKCTCSSGHGALCENECVSFFGSSCTCKODCLNNQNDSSSGEKC-IGTGTGK	3155
OY	263	VCQGPCEBGRFKNCSQECOC-----HNGTICDAATGQCHSPGYTGERCOD-ECPVGT	3155
Db	316	HODIGCSGRGGLQCKQKNCSTPGLFEFSDSNASNQAKTGGCCESGYVGRPCDEKKCOAEQ	3757
OY	316	YGVILCAETQCY--NGKCITYHSGALCJLCEAGFAEBCREARLCPRGLYIGICDKRCPHLE	3737
Db	376	YGADSKTCTCTVRETTLMCAENTFRCRKPFEYDNCEL-ACSDSYSGPNCERKAMDMN	4354
OY	374	NTSHCHPMGSGEACKPGMSGLYCNETGSPGYEAGCOOLISC--ONGADCSVTGKTCAP	4322
Db	435	HASEBNPRTGSCYVCKPRGTGKNCSEBPCPLDYGRNCAHQCCONRGVCGDADGSKCCDDR	4946
OY	433	GFKGIDCSTPCPLGTYYGICSSRCCGCKNDAYCSFVDSCTCKRAGHWGVDCSIRCPSTWG	4922
Db	495	GWTGHRCENHCRPADTFFGANCKEKRCKPFGICDPIITGECTCPALDQANDIDICPEGSYG	5544
OY	493	FGCNTTQOCLLNGAGCNTLDGTCTCAPRG-----RGRKCEL--PCOD-----	5320
Db	555	PGCKLHCACVH-GKCDKCTGTECTCOPGFSGFSDCSTTCSKGYGSECELSCPCSASCSKO	6133
OY	532	-----GTYLNCBAERCD-----	5433
Db	614	TGKCLCPLGITGVASODKCDPNTFGCLQDETVPSPCASTDPKNGVLSLCPRSSGIIHCE	6733
OY	544	-----CSHAD--GCHPTTGHCRLCPMSGVNCHDSVCAEGRMKRPMSCIPC	585
Db	674	HNCPRAGSYGDCSOOVYSCADBDGDDPPTTGEDICERGVNHGKTCSEKPCPDYGVGACALDC	732

RESULT 11		
Q8T3A7		
ID	Q8T3A7	PRELIMINARY: PRT: 1070 AA.
AC	Q8T3A7.	
DC	01-JUN-2002 (TrEMBLrel. 21, Created)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Y47H9C.4b protein.	
GN	Y47H9C.4b.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;	

OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.R.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: AL032657; CAD27614.1;
 SQ SEQUENCE 1070 AA; 114180 MW; 75254D0DD5643A5 CRC64.

Query Match 35.7%; Score 1284.5; DB 5; Length 1070;
 Best Local Similarity 34.2%; Pred. No. 6.5e-116;
 Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OY 21 GASPILNLEDPNVGSHMSESVTVQESYHPHFDIYTT-----SCDILNMFECTRHR 73
 DB 35 GTTTP--QGDRVCT-----VKTIVDDY--ELKKVIHTVYVNDTEQCLNPLTGFQC----- 80
 OY 74 VSYRTAYRHGKTYRRK-----SOCCPFYSEGE-MCVPHCADKCVHGRCIAPNTC 124
 DB 81 ----TVERKGKASYQRLVKREKYVKCCDGYQTKHFLCPDCNPCKKCKIEPKC 136
 OY 125 OCEPWGNGTNCSSACDGDHMGPHCTSRQCKNGALCNPTTGACHCACAGFRGRCE----- 179
 DB 137 ECDPQYGGKAYCASCSVGTGAGCSKSCDCENGANCDELGTCTCTSGFQGERCKPCPD 196
 OY 180 -----DRCEGTGTGNDHQRCOCONGAT 202
 DB 197 NKMGPNVCKSCPCONGCKNKGKCVCSGDMGGEFCLNKECEGKGAECFKPCNCONGAT 256
 OY 203 CDHVTGECRCPPGYTGAFCEDLCPGKKGHPQCEORPCONGVCHHVTGECSCPSGMWGT 262
 DB 257 CDNTNGKCKICKSGYHGALCENECYVGFSGCTGCDLNNQNCSSSECC-IGMTGK 315
 OY 263 VCGQCPGEGRGKNCSSQECQ-----HNGCTCDATGQCCHCSPTGTGERCD-ECPVGT 315
 DB 316 HCDIGCSKRGFLQCKQKQCTCPGLEFSDSNASCAKATGQCCESEGYKPKCDERKCDAEQ 375
 OY 316 YGVLCATCCQY--NGKCYHVSAGALCEAGFAGRCERARLCPGLYTKCKDRCPCHE 373
 DB 376 YGADCSKTCCTCYRENTLMCAPRTGFCRCKPGFYGNCEL-ACSKDSYGNCKEKQACDMN 434
 OY 374 NTHSGHPMSGECACRPGMSGLYCNETCSPGFYGEACQOITCSC-QNGADCDSTYTGKCTCAP 432
 DB 435 HASECNPEFTGSCVCKPGRFGKNCSEPCPLDFGPNCAHQCCQNGVGCGDADGKCCQDR 494
 OY 433 GKRGIDCSTPCLGTGYGNCSSRCCKKNDVCSFYDGSCTCAAGHGVDCSTRCPSGTWG 492
 DB 495 GWTGRCHEHCADTFEFGACERKCRCPKIGCDPTGTGECOTPAAGLQGANCDIGCEGSGY 554
 OY 493 FGCNLTCCOLNAGACNTLDGTCTCAPW-----NGKCECL--PCOD----- 531
 DB 555 PGCKLHCKCVN-GKCDKTEGETCQPGFPGSDCTTCSKKGESCELSCPDSASCQ 613
 OY 532 -----GTGGLNCAERCD----- 543
 DB 614 TGKCLCLPGLTKGVSQCDKQCDPRTFGLQCEFTVTPSPCASPDKNGVCLSCPPSSGICE 673
 OY 544 -----CSHAD--GCHPTTGCHCCLPMSGVHCDSCVACGRMPCNSLPC 585
 DB 674 HNCPRAGSTGDCQYVCSADGHCADPTTGECICEPGYHGKTCSEKCPDKYGYGALDC 732

RESULT 12
 O9XMD6 PRELIMINARY; PRT; 1111 AA.
 AC O9XMD6;

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiidae; Rhabditidae;
 OC Rhabditiidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Spoat J., Wohldman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163239;
 RA Zhou Z., Hartweg E., Horvitz H.R.;
 RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 RT Engulfment in C. elegans."
 RL Cell 104:43-56(2001).
 DR EMBL: AL032657; CAA21739.1;
 DR EMBL: AF332568; AAG60061.1;
 DR HSSP: P05106; 11V2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR Pfam: PF00008; EGF_9.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR SMART: SM00180; EGF_Lam; 6.
 DR SMART: SM00261; FU; 2.
 DR SMART: SM00001; EGF-like; 5.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE: PS01186; EGF_2; 11.
 KV EGF-like domain; Glycoprotein.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F37AC008F9874 CRC64;

Query Match 35.7%; Score 1284.5; DB 5; Length 1111;
 Best Local Similarity 34.2%; Pred. No. 6.7e-116;
 Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

OY 21 GASPILNLEDPNVGSHMSESVTVQESYHPHFDIYTT-----SCDILNMFECTRHR 73
 DB 35 GTTTP--QGDRVCT-----VKTIVDDY--ELKKVIHTVYVNDTEQCLNPLTGFQC----- 80
 OY 74 VSYRTAYRHGKTYRRK-----SOCCPFYSEGE-MCVPHCADKCVHGRCIAPNTC 124
 DB 81 ----TVERKGKASYQRLVKREKYVKCCDGYQTKHFLCPDCNPCKKCKIEPKC 136
 OY 125 OCEPWGNGTNCSSACDGDHMGPHCTSRQCKNGALCNPTTGACHCACAGFRGRCE----- 179
 DB 137 ECDPQYGGKAYCASCSVGTGAGCSKSCDCENGANCDELGTCTCTSGFQGERCKPCPD 196
 OY 180 -----DRCEGTGTGNDHQRCOCONGAT 202
 DB 197 NKMGPNVCKSCPCONGCKNKGKCVCSGDMGGEFCLNKECEGKGAECFKPCNCONGAT 256
 OY 203 CDHVTGECRCPPGYTGAFCEDLCPGKKGHPQCEORPCONGVCHHVTGECSCPSGMWGT 262

```

Db 257 CDTNKCICCKSGYHAGLACNECSVGFSGCTQKCDCLNNONCDSSSGCKC-IGWTGK 315
263 VCGOPCEGRFGRKNGSOECC-----HNGCTDAATGQCHSPGTGECOD-ECOPYT 315
Db 316 HCIOIGSRGFRGLOQKONCTCPGLERSDSNASCDATGQOCESGKGFCDERKDAEO 375
316 YGVLCATCOCV--NGKCYHVSAGLCEAGFAGEREARLCEPGLYGIKCDRCPCHE 373
376 YGADCKSTCTCVNHTLMCAPNTPGRCRCKPGFYGDNCCL-AGSKDSYGRPCERQAMDMN 434
374 NTHSCHPMSEGCACRPGMSLLYCNENCSPGFYEACQOITSC-QNGADCDVTGKTCAP 432
435 HASECNPETGSCVCKPGRGKNCSEPCPLDFYPNCAHOCQCNQRGVCGDAGDKCQCDR 494
433 GFGICDSTPCPLGTGGINCSRCGCKNDAYCSPYDSCSTCKAGMHGVDCSIRPSGTWG 492
495 GMTGHCENHCPCADPTGANCERKCKPCKGIGCDPTGECTCPAGLOGANCDICPBGSTG 554
493 FGCNLTQCLNGACNTLDGTCTCARGW-----RGEKCEL--PCOD----- 531
555 PGCKLMCKCVN-GKCKMETGECTCPGPFSGDSTGSKGYESELSCPSDASCSQ 613
532 -----GTGLNCAERCD----- 543
614 TGKCLCPLGTGKVSQCDKCDPNTFGLCOETVTPSPCASTDPRKNGVCLSCPPSSGJHC 673
544 -----CSHAD--GCHPTGHCRLPGMSGVHCHDSVCAEGRMGNCSLPC 585
674 HNCPASSTYDGCQOVCSADGHCDDPTGECICEPGHGTSEKCPDGKTYGICALDC 732

```

RESULT 13

075095

PRELIMINARY: PRT: 1246 AA.

```

AC 075095: 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MEGF6 (Fragment).
GN MEGF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=98360089; PubMed=9693030;
RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
RT "Identification of high-molecular-weight proteins with multiple EGF-
RT like motifs by motif-trap screening."
RL Genomics 51:27-34(1998).
DR EMBL: AB011539; BAA32467.2; -.
FT NON_TER 1
SQ SEQUENCE 1246 AA: 130304 MW: BD8E70A0F6A2CB29 CRC64;

```

Query Match 35.6%; Score 1282.5; DB 4; Length 1246;

Best local similarity 41.8%; Pred. No. 1.2e-115;

Matches 216; Conservative 58; Mismatches 214; Indels 29; Gaps 7;

```

OY 93 OCCPGTYEGEMCVPHCAD-----KCVHRCIAIPNT--CQCEPGMGTCNCSSA 138
Db 671 RCPFG--RTGEDEADCEPEGRMGLGCEICPACQHAARCDPFGACILCPFGVSRCDV 728
OY 139 CQDDHNGPHTSRQCKNGALCNPTGACHCAAGFRGMKCEDCEBGTYGNDCHQKCCQ 198
Db 729 CPAGWGSPQRTSCANQGHCHPARGHSCSCAGWGTGFSQRCADTGHMGPPDSHPCNCS 788
OY 199 NG-ATCDHTATGCRCPGTYGATGACEDLCPGKRGPOCEGRPCONGGVCHHTVGECS 257
Db 789 AGHSCDAISGLCLCEAGYVGRPCEDQCFQGHGFGCEBOLCQCHGAACDHSVAGACTCPA 848

```

```

OY 258 GMMGTVCOPCPBGRGRKNCSECCCHNGTCDATGQCHSPGTGTERCQDECPVYTG 317
Db 849 GMTGTCENHACPAFGFGLDORSAACNTAGAACAVANGSCLCPARGRPRCAEKCLPRDR 908
OY 318 VILAEFTCCYNGKCYHVSAGLCEAGFAGEREARLCEPGLYGIKCDRCPCHE 377
Db 909 AGCRHSGGCLNGGLCDPHHGRCLCPAGWTDKQOSP-CLRGWGEACARCSG--PPAA 965
OY 378 CHPMSEGCACRPGMSLLYCNENCSPGFYEACQOITSCQNGADCDVTGKTCAPFGKI 437
Db 966 CHHTVACNCRCPFGFTSGCEGCGPPRGYBGCQLGLGNGSCDATTACRCPTFLGT 1025
OY 438 DCSTPCPLGTGGINCSRCGCKNDAYCSPYDSCSTCKAGMHGVDCSIRPSGTWG 497
Db 1026 DCNLTCPGFRGPNCTHVCGCGGGAACDPYTGCTCLPPRAGVGRCBGRCPQNFVGC 1085
OY 498 TCQCLNGACNTLDGTCTCARGW-----RGEKCEL--PCOD----- 531
Db 1086 TCSGRNGGLCHASNGSCSGIGTGRHCELACPPRGYGAACHLECSCHNNSTCEPATGTC 1145
OY 558 RCLPGW-----SGVHODSVCAEGRMGNCSLPC 585
Db 1146 RCGPFGYGAACENHRSGATCNLDRCRGRQFSPSCLHC 1182

```

RESULT 14

09W0A0

PRELIMINARY: PRT: 434 AA.

```

AC 09W0A0: 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CG18172 protein.
GN CG18172.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Chame M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Adayanil A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Danke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz G.H., Ke Z., Kennan J.A., Kethum K.A.,
RA Foster C., Gorfel A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennan J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```



```

Db 161 PCPPGFYGMCKERCPEILHGKNSCDHITGEILCRITGYIGLTCHEPCPAGLYGPCKLKC 220
QY 239 PCQNGVYCHHYTGECSGSPGSMGTVCGOPCEGRFGKNCSECOCHNGTGCDATGQCHC 298
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 221 NCEHGGECNNHTGOCCLPMTGSCNENESCPDTYGGCAQRCRCVHHKCYCKRADGMCIC 280
QY 299 SPGYTGERCODECPVGYVLCALCTCQVNGKCYHVSACLCBAGFAGERCEARLCPGEG 358
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 281 ETGWSGTRCDEVCPEGFYGEHCMTACPSANFQCHAHGCVCRSGYTGDNCDLIASQR 340
QY 359 LYG-----IKDKRCPCHELTHSCHPMGGE 384
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 341 IADQSENSRASVALTLVLTFLACILFAVEFYRRRVSNLKTETIAHVTHTDTP----- 396
QY 385 CACKPGWSGLYCNETCSPGFYG-EACQOICSCQNGADDCDSVTGKCTCAPGFKGIDCSTPC 443
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 397 -----PSMPP---NNFNPNPYGMAETRLPLNNMRSKMNNPDQSTYMTDY-GDDCNASG 448
QY 444 PLGTGGINCSSRCGCKN 460
    |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 449 RVGSYSINYNHDLTLTKN 465

```

Search completed: May 9, 2003, 11:57:14
 Job time : 40.9328 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:39:55 ; Search time 64.0672 Seconds

(without alignments)
3666.364 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVLINSLSCSLFICLLCHMT.....SSPKQDSGSSSSSSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp.Archea:*
3: sp.Bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.virus:*
14: sp.vertibrate:*
15: sp.unclassified:*
16: sp.virus:*
17: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6744	100.0	1140	4	Q96K67
2	3769	55.9	969	4	Q96K66
3	3468	51.4	567	4	Q8WUL3
4	2668	39.6	1034	11	Q8VHL7
5	2667	39.5	1034	11	Q8VHL7
6	2292.5	34.0	747	11	Q8VHL7
7	1958	29.0	1574	11	Q8VHL7
8	1931	28.6	1664	5	Q9TVQ2
9	1823.5	27.0	1070	5	Q8T3A7
10	1805.5	26.8	1111	5	Q9XWD6
11	1804.5	26.8	1246	4	Q75095
12	1804	26.7	1045	5	Q8T3A6
13	1309	19.4	220	11	Q63404
14	1252	18.6	434	5	Q9W0A0
15	1024.5	15.2	2447	13	Q13149
16	1006	14.9	2653	5	Q25253

17	988.5	14.7	2524	5	Q9PA5	Q9PA5 branchiost
18	964.5	14.3	2531	5	Q16004	Q16004 lytechinus
19	920.5	13.6	4006	11	Q35452	Q35452 mus musculu
20	916.5	13.6	491	4	Q8TEK2	Q8TEK2 mus saplen
21	916.5	13.6	4288	4	Q9NPK9	Q9NPK9 homo saplen
22	915.5	13.6	2352	5	Q61240	Q61240 halocynthia
23	903.5	13.4	4114	11	Q54796	Q54796 mus musculu
24	881.5	13.1	4135	6	Q18977	Q18977 bos taurus
25	863	12.8	713	5	Q962M9	Q962M9 podocoryne
26	832.5	12.3	752	13	Q42374	Q42374 brachydanto
27	790.5	11.7	1214	13	Q90XD2	Q90XD2 xenopus lae
28	782	11.6	1254	13	Q9YXU2	Q9YXU2 brachydanto
29	782	11.6	1254	13	Q90X56	Q90X56 brachydanto
30	779	11.6	1193	13	Q90819	Q90819 gallus gall
31	765	11.3	1212	13	Q42347	Q42347 gallus gall
32	763.5	11.3	2146	5	Q9VC97	Q9VC97 drosophila
33	762	11.3	1216	13	Q90X55	Q90X55 brachydanto
34	748.5	11.1	2019	11	Q64706	Q64706 mus musculu
35	741	11.0	594	5	Q9W0A1	Q9W0A1 drosophila
36	740	11.0	594	5	Q9Y151	Q9Y151 drosophila
37	735.5	10.9	3695	4	Q8TDE8	Q8TDE8 homo saplen
38	723	10.7	3367	5	Q9XZC9	Q9XZC9 drosophila
39	720.5	10.7	1532	13	Q90994	Q90994 gallus gall
40	720.5	10.7	1714	13	Q90995	Q90995 gallus gall
41	711.5	10.6	2656	5	Q9GND3	Q9GND3 paracentrot
42	709.5	10.5	1810	13	Q90824	Q90824 gallus gall
43	707.5	10.5	1799	11	Q8RYO0	Q8RYO0 mus musculu
44	700.5	10.4	1792	13	Q57484	Q57484 gallus gall
45	696.5	10.3	3319	5	Q9VJ75	Q9VJ75 drosophila

ALIGNMENTS

RESULT 1
ID Q96K67 PRELIMINARY; PRT; 1140 AA.

AC Q96K67;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MEGF10 protein (K1A1780).
GN MEGF10.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HIPPOCAMPUS;
RX MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 8:85-95(2001).
DR EMBL: AB058676; BAB47409.1; -
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF_15.
DR PROSITE: PS00022; EGF_1; UNKNOWN_17.
DR PROSITE: PS01186; EGF_2; UNKNOWN_17.
SQ SEQUENCE 1140 AA; 122204 MM; 45B2FA239423895A CRC64;

Query Match 100.0%; Score 6744; DB 4; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLINSLSCSLFICLLCHMTGTAAPLNEPNCVSHESVYVQESYPPFDIYTSC 60
DB 1 MVLINSLSCSLFICLLCHMTGTAAPLNEPNCVSHESVYVQESYPPFDIYTSC 60
QY 61 TDLINFKCTRHRYSTAYRNGEKTMYRRKSQCCPFYEGSCVPHCADKCHGRGIA 120
|||||

```
Db 61 TDILMFKCTRRHVSRYTRAYRHGEKTMTRKSOCCPFYEGSEGMVPHCADKRCVHGRCIA 120
Qy 121 PNTCCPEPMGGTNCSSACDDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
Db 121 PNTCCPEPMGGTNCSSACDDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
Qy 181 REBOGTGNDCHORCOQONGATCDHVTGBCRCPRGYTGAFCEDLCPRGHRGPOCEGRCP 240
Db 181 REBOGTGNDCHORCOQONGATCDHVTGBCRCPRGYTGAFCEDLCPRGHRGPOCEGRCP 240
Qy 241 ONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 300
Db 241 ONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 300
Qy 301 GTTGERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPGLY 360
Db 301 GTTGERCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPGLY 360
Qy 361 GIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEACQOITCSQNGAD 420
Db 361 GIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEACQOITCSQNGAD 420
Qy 421 CSVTGKCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 480
Db 421 CSVTGKCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 480
Qy 481 DCSIRPCSGTWGFGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 540
Db 481 DCSIRPCSGTWGFGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 540
Qy 541 RDCDSHADCCHPTGHCRLPBGSHYHCDYVCAEGMGNCSIPCTCKNGASCPDGTIC 600
Db 541 RDCDSHADCCHPTGHCRLPBGSHYHCDYVCAEGMGNCSIPCTCKNGASCPDGTIC 600
Qy 601 ECAPGRGTTCORICSPGEGYHRCSTQCPCYHSSGPRHHITGCLDCLRGFTGALCNEYC 660
Db 601 ECAPGRGTTCORICSPGEGYHRCSTQCPCYHSSGPRHHITGCLDCLRGFTGALCNEYC 660
Qy 661 PSGRREGKNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 720
Db 661 PSGRREGKNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 720
Qy 721 GAFCSAYDECKCTPGMTLCTGTCORPLGTFYGDCLLTCQONGACDHIISGCTCTRTGF 780
Db 721 GAFCSAYDECKCTPGMTLCTGTCORPLGTFYGDCLLTCQONGACDHIISGCTCTRTGF 780
Qy 781 MGRHCEQKCPSGTYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 840
Db 781 MGRHCEQKCPSGTYGCGHQICDCLNNSTCDHITGTCYCSPGMKGARCDQAGYIYGNLN 840
Qy 841 SLSTRSTALPADSYOIGAIAGIILVVLFLALFTIYRHKOKGESSMPAVTYTPAMR 900
Db 841 SLSTRSTALPADSYOIGAIAGIILVVLFLALFTIYRHKOKGESSMPAVTYTPAMR 900
Qy 901 VYNADVTISGTLPHSNGNANSHYFNPSYHTLTGCATSPHYNNRBMVTKSKNOLTV 960
Db 901 VYNADVTISGTLPHSNGNANSHYFNPSYHTLTGCATSPHYNNRBMVTKSKNOLTV 960
Qy 961 NLKNVNPGRGVPDCTGTLPADMKHGYLNLGAFGLDRSYMGKSLKJLKGKSEYNSN 1020
Db 961 NLKNVNPGRGVPDCTGTLPADMKHGYLNLGAFGLDRSYMGKSLKJLKGKSEYNSN 1020
Qy 1021 CSLSSSENRYATIKDPVLIPIKSSBEGYEMKSPARBDSPYAEINNSTANRNVYEVEPT 1080
Db 1021 CSLSSSENRYATIKDPVLIPIKSSBEGYEMKSPARBDSPYAEINNSTANRNVYEVEPT 1080
Qy 1081 VSVVGVGFNSNGRLSDPDLPRKNSHIPCHYDLPYRDSSSSPKODSGSSSSSSSE 1140
Db 1081 VSVVGVGFNSNGRLSDPDLPRKNSHIPCHYDLPYRDSSSSPKODSGSSSSSSSE 1140
```

```
Id Q96KG6 PRELIMINARY; PRT; 969 AA.
AC Q96KG6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE MEKFI1 protein (K1MAL781).
GN MEKFI1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=BRIN;
RC MEDLINE=21245130; PubMed=11347906;
RA Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 8:85-95(2001).
DR EMBL: AB058677; BAB47410.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001281; Rleske.
DR Pfam: PF00008; EGF_15.
DR PROSITE: PS00022; EGF_1; UNKNOWN_17.
DR PROSITE: PS01186; EGF_2; UNKNOWN_17.
DR PROSITE: PS00200; RIESKE_2; UNKNOWN_1.
SQ SEQUENCE 969 AA; 101600 MW; 56DDJFEFL39C8209 CRC64;

Query Match 55.9%; Score 3769; DB 4; Length 969;
Best Local Similarity 58.6%; Pred. No. 1,2e-296;
Matches 600; Conservative 126; Mismatches 208; Indels 90; Gaps 4;

Qy 109 CADKCYHGRICAPNTCCPEPMGGTNCSSACDDHMGPHCTSRCCCKNGALCNPTTGACH 168
Db 109 CADKCYHGRICAPNTCCPEPMGGTNCSSACDDHMGPHCTSRCCCKNGALCNPTTGACH 168
Qy 28 CTEBCHGKCYSPDTCHEGRCMGGRPCSCGSDHMGPHCTSRCCCKNGALCNPTTGACH 87
Db 28 CTEBCHGKCYSPDTCHEGRCMGGRPCSCGSDHMGPHCTSRCCCKNGALCNPTTGACH 87
Qy 169 CAAGFRGMRCEELCAPGTHGKGCQLPCQCRHGASCDPRAGELCAPGYGVYCEELCPG 147
Db 169 CAAGFRGMRCEELCAPGTHGKGCQLPCQCRHGASCDPRAGELCAPGYGVYCEELCPG 147
Qy 229 KHGRPOEORCPONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 288
Db 229 KHGRPOEORCPONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 288
Qy 148 SHGARELCRCPONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 207
Db 148 SHGARELCRCPONGGVCHHTGECSCPSGSMGTVCQGPCEGRGRKNCSECCCHNGCTCDATGQCHCSP 207
Qy 289 CDAATGCHCSPYTGECRCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPGLY 348
Db 289 CDAATGCHCSPYTGECRCODECEPVGTGVLCAETCCQVNGKCYHVSAGALCEAGFAGERCLEARLCEPGLY 348
Qy 208 CDHVTGBCRCPRGYTGAFCEDLCPRGHRGPOCEGRCP 267
Db 208 CDHVTGBCRCPRGYTGAFCEDLCPRGHRGPOCEGRCP 267
Qy 349 KCEARLCPEGLYGIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEA 408
Db 349 KCEARLCPEGLYGIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEA 408
Qy 268 RCOERLCPEGLYGIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEA 327
Db 268 RCOERLCPEGLYGIKCDKRCPCHEBTHSCHPMSEGCACRPMGSLYCNETCSPGFYGEA 327
Qy 409 COOICSCONGADSDYTGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 468
Db 409 COOICSCONGADSDYTGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 468
Qy 328 COLPCTCQNGACDHSITGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 387
Db 328 COLPCTCQNGACDHSITGCTCAPGFKGIDSTPCPLGTGYNCSRCGCKNDVACSPYDSCCTKAGMHV 387
Qy 469 GSCTCAGHAGVDCSIRCPSTGTFEGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 528
Db 469 GSCTCAGHAGVDCSIRCPSTGTFEGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 528
Qy 388 GSCTCAGHAGVDCSIRCPSTGTFEGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 447
Db 388 GSCTCAGHAGVDCSIRCPSTGTFEGCNLTCCCLNGAGCNTLDGTCTCAPGMRGECBLPCODGTGGLNCAE 447
Qy 529 CODGTGGLNCAEAGFAGERCLEARLCEPGLY 588
Db 529 CODGTGGLNCAEAGFAGERCLEARLCEPGLY 588
Qy 448 CPDGTGGLNCAEAGFAGERCLEARLCEPGLY 507
Db 448 CPDGTGGLNCAEAGFAGERCLEARLCEPGLY 507
Qy 589 NGASCPDGTGGLNCAEAGFAGERCLEARLCEPGLY 648
Db 589 NGASCPDGTGGLNCAEAGFAGERCLEARLCEPGLY 648
Qy 508 NGASCPDGTGGLNCAEAGFAGERCLEARLCEPGLY 567
Db 508 NGASCPDGTGGLNCAEAGFAGERCLEARLCEPGLY 567
Qy 649 PGFTGALCNEVCPSGRFRGNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 708
Db 649 PGFTGALCNEVCPSGRFRGNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 708
Qy 568 PGFTGALCNEVCPSGRFRGNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 627
Db 568 PGFTGALCNEVCPSGRFRGNCAGICTCTNNCTNPIDRSCOCYPMGIGSDCSQPCPAHMGPNCTIHTCNCHN 627
```

```

OY 709 GNCITTCNCHNGAFCSAYDGECKTPEWTGLYCTORCPGLGYGKCALICCONGACDD 768
DB 628 GRACFHACCHNGASASADGACHCTPGMTGLFCTQRCAPAFGKDCGVCCQONASCD 687
OY 769 HISSGCTCRFGMRGRHCEOKCPSTGYGCRQICDCLNNSTCDHTTCTCSPGMKGAR 828
DB 688 HISSGCTCRFGMRGRHCEOKCPSTGYGCRQICDCLNNSTCDHTTCTCSPGMKGAR 828
OY 829 DDAGYIVNLSLSTSTALPADSVOIGALINGIIVLVYFLALFYHAKOKGKS 888
DB 748 DDA-ALMELNLPYTKISALAEKRSVGAIVGIMLLFIVVLGLFAMHRRROEKER 806
OY 889 SM-PATYTPARAVYNADITISGLPHSNGNANSHYFNPSYHTLLOCATSPHYNNRDR 947
DB 807 DLAPRYSYTPAMRMTSTDYSL- 828
OY 948 MIVYSKNNQLVNLKNVPRGRGVGDCTGLPADMKHGYLNLGAFGLDRS- 1003
DB 829 -----GACGMDRRQNTYIM 842
OY 1004 GKSLDGLKNSSEYNSNCSLSSSENPYATIKDPVILPKSBCGYEMKSPARRDSYAE 1063
DB 843 DKGFDDYKESVCSSTSCSLSSSENPYATIKDPVILPKSBCGYEMKSPARRDSYAE 902
OY 1064 INNSTSANRNYEVEPTVSVGVFNSNGRLSODPYDLPKNSHIPCHYDLFVRDSSSP 1123
DB 903 VSLSTSNKNITEVEPTVSVGVFNSNGRLSODPYDLPKNSHIPCHYDLFVRDSSSP 962
OY 1124 KOED 1127
DB 963 PSQD 966

```

RESULT 3

```

O8WUL3 PRELIMINARY: PRT; 567 AA.
AC O8WUL3;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to Mesf10 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC020198; AAH20198.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF002049; Laminin_EGF.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_11.
DR SMART: SM00180; EGF_Lam; 9.
DR PROSITE: PS00022; EGF_1; UNKNOWN_10.
DR PROSITE: PS01186; EGF_2; UNKNOWN_10.
SQ SEQUENCE 567 AA; 60797 MW; CF2FB8CDEB7CF627 CRC64;

```

Query Match 51.4%; Score 3468; DB 4; Length 567;
 Best Local Similarity 99.8%; Pred. No. 1.5e-272;
 Matches 565; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 NVISINSLSTFICLLCHNIGTASPLNEDPNVCSHWESYVTVQESYPHPDQIYITSC 60
DB 1 NVISINSLSTFICLLCHNIGTASPLNEDPNVCSHWESYVTVQESYPHPDQIYITSC 60
OY 61 TDLNMFCTRRHVSRYFAIRGEGTKYTRRKSCCPGFESGEMCVPHCADKCVHGRCIA 120
DB 61 TDLNMFCTRRHVSRYFAIRGEGTKYTRRKSCCPGFESGEMCVPHCADKCVHGRCIA 120

```

```

OY 121 PNTCCCEPBGWGTNCSACDGDHMBPCHTSRCCCKNGALCNPTGACHCAAGFGRMCD 180
DB 121 PNTCCCEPBGWGTNCSACDGDHMBPCHTSRCCCKNGALCNPTGACHCAAGFGRMCD 180
OY 181 RCEQGTGNDCHQRCOQONCATCDHYTGRCRCPPTGTAFCEDLCPGKHGPOCEORPC 240
DB 181 RCEQGTGNDCHQRCOQONCATCDHYTGRCRCPPTGTAFCEDLCPGKHGPOCEORPC 240
OY 241 QNGVYCHHYTGSCSPSGMGTVCQPCPEGRFKNKSQCECCHNGGTDAATGQCHSP 300
DB 241 QNGVYCHHYTGSCSPSGMGTVCQPCPEGRFKNKSQCECCHNGGTDAATGQCHSP 300
OY 301 GYTGRCODECPVGTGYVLCAETCCQVNGCKYHNSGACLCBAGEGCEARLPEGLX 360
DB 301 GYTGRCODECPVGTGYVLCAETCCQVNGCKYHNSGACLCBAGEGCEARLPEGLX 360
OY 361 GIKCKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETGSPGYGEACQITSCONGAD 420
DB 361 GIKCKRCPCHELNTHSCHPMSGECACRPGMSGLYCNETGSPGYGEACQITSCONGAD 420
OY 421 CDSVTKGTCAAGFGICDSTPCPLGTGYNSSRCCGKNDAVCSBPVDSCTCKAGMHV 480
DB 421 CDSVTKGTCAAGFGICDSTPCPLGTGYNSSRCCGKNDAVCSBPVDSCTCKAGMHV 480
OY 481 DCSICPSTGWFNCNLFQCCLNGACNTLDGTCTCAPGMRKEKELPCQDGTGLNCAE 540
DB 481 DCSICPSTGWFNCNLFQCCLNGACNTLDGTCTCAPGMRKEKELPCQDGTGLNCAE 540
OY 541 RCDGSHADGCHPTTGHCRCLPGMSGV 566
DB 541 RCDGSHADGCHPTTGHCRCLPGMSGV 566

```

RESULT 4

```

O8VHL7 PRELIMINARY: PRT; 1034 AA.
AC O8VHL7;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Jedi protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=TESTIS;
RA Kriytsov A.V., Zinovyeva M.V., Hendrikx J., Vleser J.W.M.,
RA Belyavsky A.V.;
RL "Jedi is a novel DSL and EGF-like repeat motif-containing protein
expressed on non-differentiated hematopoietic cells."
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF444274; AAL38571.1; -.
DR InterPro: IPR000561; EGF-like.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00181; EGF_15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 1034 AA; 5514E5166AE01111 CRC64;

```

Query Match 39.6%; Score 2668; DB 11; Length 1034;
 Best Local Similarity 42.7%; Pred. No. 2e-207;
 Matches 493; Conservative 110; Mismatches 379; Indels 172; Gaps 16;

```

OY 14 LILCHWIGTASPLNEDPNVCSHWESYVTVQESYPHPDQIYITSCDILNM---FCT 70
DB 7 LILALGLRTGLTNSNDPNVCSHWESYVTVQESYPHPDQIYITSCDILNM---FCT 70
OY 71 RHRVSRYFAIRGEGTKYTRRKSCCPGFESGEMCVPHCADKCVHGRCIA 130

```

```

Db 65 QPTVYRTVYKQVYKMDSRPRLQCCRGYESRGACVPLCAQECVHGRCVARNOCQCAFGW 124
OY 131 GGTNCSACDGDHNGPHTSRQCKNGALNPITGACIACAGFGRNRCEDRCDEGTGND 190
125 RGGDCSSSACAGMMGPQCDKFCCHGNNSSCDPKSACCPGSLQCPNCLQCPRGHGPRA 184
OY 191 CHORCOONGATCDHVTGECRCPRGYTGAFCEIDLCPGKHGRPOCEQRCPONGGVCHHTV 250
185 CQFPCQCY - GASCDPQDAGCFPCPRGAPSPCNVPCSOCTDFCFPRYPCQNGVPGQSQ 243
OY 251 GEGSCPGMGTGCGOPREGRFGKNCSEOCQCHNGGCDATATQCHSPRYTERCODE 310
244 GSCCPRMGMVICSILPCBGFHNPCTQECRCHNGGLCDFTQCHCAPGYIDRCQOE 303
OY 311 CPVGTGYLCAETQCVNGKCYHVSGACLCCEAGFAGEBARCLPREGYLKCDKRCPC 370
304 CPVGRFGQDCAETDCAPGARCFRANAGACLEHFTGDRCTERICLPGRYLSQDEPCTC 363
OY 371 HLENTHSGHPSGECACKPFGHSLYCNETSPGFYGEACQOICSGQNGADCVTGKCTC 430
364 DPEHSLSCHPMHGECSCOPGWAHGLHCNESCPDTHGPGCOHCLLHGGLADSLGCLRC 423
OY 431 APFGKIDCSTPCPLGTGYNCSRCGCKNDVCSPYDGSCTCKAGHNGVDCSTRCSGT 490
424 APGTGPHCANLCPPTYGICSSRSCENAIACSPIDGTCTCEGWRGNCVPCPLGT 483
OY 491 WGFNCNLTCQCLNGACNTLDGTCTCARGWGERKCELPDODGTGYLNCABERDCSHADG 550
484 WGFNCNASCCCAHNDGVCSPGTGACTCTPGMHGACQLPCPRGQRBEGCASVCDCHSDG 543
OY 551 HPTGHCRCLPBGHGVHCDSCAEGRWGPNCLPCTCKNAGNSCPDDBGICCAAGFGTT 610
544 DPHNGQRCQAGMGTGCHLPCBGFHNGANCNTCTCKNGTCSSENCNCACAFGRPS 603
OY 611 CORICSPGFYGHRCSTQPCQVHSSGPHHTGLDCLPFTGALCNEVCSGRFNKA 670
604 CQRCPPRGGRKRCVQ----- 619
OY 671 GICCTNN-GTCNPIRDSQCYPMIGSDCSQPCRAHMGPNCTIHTCNHNGAFCSAYD 729
620 --CNCNNHSSCHPSDCTSCLAGMTGPDSEACPRGMGLKCSQLCQCHNGGCHPDG 677
OY 730 ECRCTPMTGLYCQRCPLRFGKDCALICQCCONGADCDHISGCTCTRTGMRHNEQKC 789
678 SCICTPMTGPNCLGSCPRPFYGVCSQLCQCDG----- 712
OY 790 PSGTYGCRQICDCLNNSTCDHTTCTCSPGKRGARCDQAGVLIIGNLSLSTAL 849
713 -----EKNCHPOTGACVCPRGHSGADK-----MGSQESTIMPTS- 747
OY 850 PADSTQICAGIILVVLFLALFIYRHQKGRHSSMPATYTPAMRVVNADYTIS 909
748 PVTNHSAGAVIAGLVALTALFALFIRYOMQKREHHLAVASTG-RLDSDDYMP 806
OY 910 GILPHSGNANSHTFNPSTHTLQCATSPHVNNRBRMTYTKSNQOLVNLKNVPGK 969
807 DVSP-----SYSHYISNPSTHTLSQSPNPPRN-----KVPGSQLEVSSQAPRPS 853
OY 970 RGVGVDCSTGLPADMKHGGYLINEAGFGLDRSY-----MGKS 1006
854 RAHGREHNTYLPADMKRRRPHRGASHLDRSTSCSHHNGOPFCGHKRPISREGIGAS 913
OY 1007 LKDLGNSEYSSNCSLSSENPRYATIKDPVLLPKSSEGVEMKSPARBDSPYAIINN 1066
914 VMSL-----SSENPYATIRDLPSLGRGREGYEMKPPSVSPRQSLH- 958
OY 1067 STANANRYVER-----TVSVQGVFSSNNGRLSQDB-----YDLPKNSHIFCHY 1111
959 --LRDRQORQLQPORDSGTYEOPSPLSHNEESLSTPRLPGLPQGYDSPKNSHIGHY 1016
OY 1112 DLLPVDRSSSPKQ 1125
1111 1111 1111

```

```

Db 1017 DLLPVHRPPSPSR 1030
RESULT 5
OBV1K5 PRELIMINARY; PRT; 1034 AA.
AC OBV1K5;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-JUN-2002 (Tremblrel. 20, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MGEF12.
GN MGEF12.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=LIVER;
RA Ivanova N.B., Lemischka I.R.;
RT "The global gene expression profiling of the hematopoietic stem
cell."
RT cell.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF440279; AL33583.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PR00011; EGF/LAMININ.
DR SMART: SM00181; EGF_15.
DR SMART: SM00180; EGF_Lam; 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_13.
DR PROSITE: PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 1034 AA; 110580 MW; 714E5016848E4EAC CRC64;

Query Match 39.5%; Score 2667; DB 11; Length 1034;
Best Local Similarity 42.8%; Pred. No. 2.4e-207;
Matches 494; Conservative 110; Mismatches 378; Indels 172; Gaps 17;

OY 14 LLLCHMIGTASPLNLEDPNVCSSHWEYSVTVOESYPRPFQIYVTSCTDILNM---FKCT 70
7 LLLALGRLGTLNSNDPNCTFWESTTTTKESHLLPFLRAESCH--RPNEDPHTCA 64
Db 71 RHRYSYTAIRHGEKTYRRKSSQCCPGFYSGEMCVPHCADKCVHGKCIAPNTQCEPGW 130
65 QPTVYRTVYKQVYKMDSRPRLQCCRGYESRGACVPLCAQECVHGRCVARNOCQCAFGW 124
OY 131 GGTNCSACDGDHNGPHTSRQCKNGALNPITGACIACAGFGRNRCEDRCDEGTGND 190
125 RGGDCSSSACAGMMGPQCDKFCCHGNNSSCDPKSACCPGSLQCPNCLQCPRGHGPRA 184
OY 191 CHORCOONGATCDHVTGECRCPRGYTGAFCEIDLCPGKHGRPOCEQRCPONGGVCHHTV 250
185 CQFPCQCY - GASCDPQDAGCFPCPRGAPSPCNVPCSOCTDFCFPRYPCQNGVPGQSQ 243
OY 251 GEGSCPGMGTGCGOPREGRFGKNCSEOCQCHNGGCDATATQCHSPRYTERCODE 310
244 GSCCPRMGMVICSILPCBGFHNPCTQECRCHNGGLCDFTQCHCAPGYIDRCQOE 303
OY 311 CPVGTGYLCAETQCVNGKCYHVSGACLCCEAGFAGEBARCLPREGYLKCDKRCPC 370
304 CPVGRFGQDCAETDCAPGARCFRANAGACLEHFTGDRCTERICLPGRYLSQDEPCTC 363
OY 371 HLENTHSGHPSGECACKPFGHSLYCNETSPGFYGEACQOICSGQNGADCVTGKCTC 430
364 DPEHSLSCHPMHGECSCOPGWAHGLHCNESCPDTHGPGCOHCLLHGGLADSLGCLRC 423
OY 431 APFGKIDCSTPCPLGTGYNCSRCGCKNDVCSPYDGSCTCKAGHNGVDCSTRCSGT 490
424 APGTGPHCANLCPPTYGICSSRSCENAIACSPIDGTCTCEGWRGNCVPCPLGT 483
OY 491 WGFNCNLTCQCLNGACNTLDGTCTCARGWGERKCELPDODGTGYLNCABERDCSHADG 550
484 WGFNCNASCCCAHNDGVCSPGTGACTCTPGMHGACQLPCPRGQRBEGCASVCDCHSDG 543

```

```

OY 551 HPTTGHCRCLPFGMSGVHCDVCAEGRMGPNCSLPCTCKNGKASCSPDDGICECAPGERGT 610
DB 544 DPHVQOCROQAGMGTGRLPCPEBGFMGANCSNTCTCKNGGTVCBENGNCVAPGRGFS 603
OY 611 CORICSPGFYGRCSQTCPOCVHSSGPCHHITGLDCCLPGFTGALCNEVCPSGRGKNC 670
DB 604 CORPCPPRGYGRKCYO----- 619
OY 671 GTCCTNN-GTCNPIIDRSQCYPGMIGSDCSPCPRAHMGPCITTCNCHNGAFCSAYDG 729
DB 620 --CKNNHNSCHPDSGTCSCLAGMTGPDSCSPRGHMGKLSQLOCHHGCTHPDGG 677
OY 720 ECKTPTGWTGLCTGRCPLGFYGCDCALICQONGADCDHISGQCTCRGTGFGRCCEQK 789
DB 678 SCICTPGMTGPNCLGECPPRMFGVNCSQLCCDLGEMC-----HPE--- 718
OY 790 PSGTYGRCROICDCLNNSYCDHITGTCYCPGMKARGADQAGVLIIVGNLNSLSRSTAL 849
DB 719 -----TGACVCPRGHSGADCK-----MGQSSEFTIMPIS- 747
OY 850 PADSTQICAGIITLLVLEFLALFTYTRHKQKGSMPAVYTTTPAMRYVNDYTS 909
DB 748 PTHNSLAVIGIAVLGTLVALLALFTGYROMQKQKEHHLAVASTG-RDGSDDYMP 806
*OY 910 GTPHNSNGNANSHYFTNPSYHTLQCATSPVNNRDMRTVTKSKNNOLFVNLKXVNGK 969
DB 807 DVPSP-----SYSHYSNPSYHTLSQCSPPPPN-----KVPQSLEFVSQAPERPS 853
OY 970 RGPVDCGTGLPADMKHGGLNELGAFGLDRSY-----MGKS 1006
DB 854 RAHGRENHVTLADMKHREPERHGASHLDRSYSCSYSHRNGPGRCHKGPISERGLAS 913
OY 1007 LKDLKNEBYNSNCSLSSENPATIKDPVYLIRKSSCCGYEAKSPARRDSPLAEIINN 1066
DB 914 VMSL-----SSENPVATIRDLPSLPGEPRESGYEAKGPPSVSPQSLH- 958
OY 1067 STSANRNYEVEP-----TVSYVQGVFSNNGRLSODP-----YDLFNSHIPCXY 1111
DB 959 --LRBROQRQIQPDHDSGVEBPSPLSHNEBSLSTPPLPCLPGRGHDSPPKNSHHPGHY 1016
OY 1112 DLLPYRDSSSSPKQ 1125
DB 1017 DLLPYRHPPSPPSR 1030

RESULT 6
O8VHF4 PRELIMINARY; PRT; 747 AA.
AC O8VHF4;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Jcd1-736 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL; TISSUE=TESTIS;
RA Kriyavsov A.V.; Zinovyeva M.V.; Hendriks J.; Vlasov J.W.M.,
Belyavsky A.V.;
RT "Jcd1 is a novel DSI and EGF-like repeat motif-containing protein
expressed on non-differentiated hematopoietic cells."
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF461685; AAL6380.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF_9.
DR PRINTS: PRO0011; EGFLAMININ.
DR SMART; SM00181; EGF_15.
DR SMART; SM00180; EGF_Lam; 14.

```

```

DR PROSITE; PS00022; EGF_1; UNKNOWN_13.
DR PROSITE; PS01186; EGF_2; UNKNOWN_12.
SQ SEQUENCE 747 AA; 78972 MW; F825FEF384D4736A CRC64;

Query Match 34.0%; Score 2292.5; DB 11; Length 747;
Best Local Similarity 48.5%; Pred. No. 3,5e-177;
Matches 377; Conservative 75; Mismatches 272; Indels 53; Gaps 5;

OY 14 ILLCHWIGTASPLNLEDPVNCVSHMSYSVTQGESVPHRPDQIYYNSCDILW---FNGT 70
DB 7 ILLALGLRLTGLNSNDPVCYFWESFTTTRKSHLRPSLPAESCH--RPMEDPHICA 64
OY 71 RHRYSPRTAYRAGEKTYRRKSSQCCPGYSESEMCPHACDCKVAGRCIAPRTCCCEPM 130
DB 65 QPTVYRIVYRKYVAKMDSRPLQCCRGYTESGACVPLCAQDCHVIGRCVAPRQCCARGM 124
OY 131 GGTNCSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRGWRCEDECEGTGND 190
DB 125 RGDGSSBACAPQMGPOCDKFCCHGNSSCDPKSGACRCPSGLQPPNCLQPCPAGHYPA 184
OY 191 CHORCQNGATCDHYTEBRCRPPGYTGAFCEDLCPRGKRGPCQQRCPQNGVCHHYT 250
DB 185 CQFDQCY-GASCDQDADACEFCPPRAGPSCVPCPSQGTDFECPRTYPCQNGVPOGSO 243
OY 251 GECSPSGMGTGVCOPCEGRFRGKNSQECQCHNGGTCDDATGQCHSPGYTGBRCD 310
DB 244 GSCSCPFGMGVYICSLPCEBGFHGNCTQECCHNGGLCDRTGQCHCAPGTYGRDCE 303
OY 311 CPVGTGYVLCATCCOCVNGKCYHSGALCEBAGEFAGERCEARLCPGLYGIKDRKPC 370
DB 304 CPVGRFGQDCAETCCARCARCFRANAGALCEBHTGDRCTRLCPDRGYGLSCDEPCTC 363
OY 371 HLENTHSCHPMGECACRFGWGLCNCTCSGFTGEGEACQOQCSQONADCDVYGTGKTC 430
DB 364 DPEHLSCHPMHGECSQPGWAGLHCNBSCPDTHGPGCQDEICLTLHGGLCLADSGLCRC 423
OY 431 APFGKIDCSTPPLGTGTCGINSRSGCKNDVAVSPVDGSCYCKAGMHWGVDSCIRPST 490
DB 424 APGYTPGHANCLCPPTDTGINSRSCENALACSPIDGTCTCKRGWGRGNCSVPCPLGT 483
OY 491 WGFNCNLTCQCLNGACNTLDGTCTCAFRMGKEKELPCQDGTGYNLNCARDCSHADGC 550
DB 484 WGFNCNASQCAHDGVCSPQTGACTCTPGMHAHQQLPCPKQFEGGACASVDCDHSDDC 543
OY 551 HPTTGHCRCLPFGMSGVHCDVCAEGRMGPNCSLPCTCKNGKASCSPDDGICECAPGERGT 610
DB 544 DPHVQOCROQAGMGTGRLPCPEBGFMGANCSNTCTCKNGGTVCBENGNCVAPGRGFS 603
OY 611 CORICSPGFYGRCSQTCPOCVHSSGPCHHITGLDCCLPGFTGALCNEVCPSGRGKNC 670
DB 604 CORPCPPRGYGRKCYO----- 619
OY 671 GTCCTNN-GTCNPIIDRSQCYPGMIGSDCSPCPRAHMGPCITTCNCHNGAFCSAYDG 729
DB 620 --CKNNHNSCHPDSGTCSCLAGMTGPDSCSPRGHMGKLSQLOCHHGCTHPDGG 677
OY 720 ECKTPTGWTGLCTGRCPLGFYGCDCALICQONGADCDHISGQCTCRGTGFGRCCEQK 786
DB 678 SCICTPGMTGPNCLGECPPRMFGVNCSQLCCDLGEMCPQTAGVACVCPRGHSGADCK 734

RESULT 7
O88281 PRELIMINARY; PRT; 1574 AA.
AC O88281;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MEGF6.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

```



```

Db      762 TRKAGAGVGNLCDDPCPAGYGFYDCEDEKSCADVASPHKSKVCHHYTGCTCLGKXGP 841
Oy      263 VCGPCCPREGFGKNSQEQCHNGSTCDAAATGCHQCSFGYTGERCODCECPVGTGVLCAE 322
Db      842 LCDQSCAANTYGPNAHATCSGVNAGAKCDESDSCHCTPFGYATCSFVCPTRGFGIDCMQ 901
Oy      323 TCCGNGGKCYHVSGACCEAGFAGEBCEARLCPBGLGKINCDKRCPLHENTHSCHEMS 382
Db      902 LCKCNNGAICDTNSGSCCAPAGWSGKCD-KACAPGTGKDCSKCD-ADGMH-CDPSD 958
Oy      383 GECACKPMSGLYCNETSPGFYGEACQICSCQNGADCDSDVTGKCTCAGPGRK----- 435
Db      959 GECICPPKKGKHCDETDSDGLFGAGCKGICSCQNGATCDSDVTGSECEGRPGRKGKCDP 1018
Oy      436 -----GIDCSTPCPLGTG 449
Db      1019 CPDGRFGGCGNACIDCTTNDTSMNPVARDHYTGECRCPCAGWTGCDCCQTCPLGRHG 1078
Oy      450 INCSRCGCKNDVAVSPVDGSCCTKAGMHGYDCSIRCPSGTMFGPCNLTCOLNGACNT 509
Db      1079 EGRHSCCQSNAGSCDRVTGFCDCPSGFMGNKSECEPBGILWGNMCMHCLCMHGECKN 1138
Oy      510 LDGTCTCAPGMRGECCELPCCODGTGTLNCAERCDSCSHADGCHPTTGHCRCLEPGMSGVHCD 569
Db      1139 ENGDCIDGWTGSCFELCPGQGFGRMAQKNCNGASCDORKTGRCECLPGMSGEHCE 1198
Oy      570 SYCAGBRGPNCSLPCYKKNASCSPPDDGICECAPGFGTTCQIRICSGFTGHRCSOTCP 629
Db      1199 KSCVGHYGAKEETCEBENGALCDPIGHCSCQPGMRGKCNRPCLGFGYFGRHCSQSC- 1257
Oy      630 QCVHSSGPHHTTGLCDCLPFTGALCNEVCP----- 661
Db      1258 RCANKS-CDHISGCGCPKGTAGHSTELCPDGTFGESCSKCCCGENSMKDAISGKCF 1316
Oy      662 -----SGRFGKNACIGCTNNGTCNPIDRSQCCYPMIGSDSCSPCPA 706
Db      1317 CKPGHSGSDCKSGCVQGRFGPDCNQLCSCENGVCDSSTGSCVCPGYIGTKCEIACQSD 1376
Oy      707 HMGPRCIRHCNCHNCAFSAVDGECCKTPGWTGLYCTQRCPLGFGKCALICQONQAD 766
Db      1377 RFPPTCEKICENGSGTDRLTGQRCPLGFTGTCNQVCPGGRGACKCKEKCANG-H 1435
Oy      767 CDHISGQCTGRTGFMGRHCEOKPSGTGYGCGROICDCLNNSCTCHITGYCSPGMKA 826
Db      1436 CNAASGEEKCNLGTGSPCDSGSKYGLNCTLDCBCTGQARCDPVGCGCCDPPGRYGS 1495
Oy      827 RC 828
Db      1496 RC 1497

```

```

RESULT 9
Q8T3A7 PRELIMINARY; PRT: 1070 AA.
AC Q8T3A7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Y47H9C.4B.
GN
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
OC Rhabdilitidae; Peloderiine; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;

```

```

RT "Genome sequence of the nematode C. elegans: A platform for
RT Investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; AL032657; CAD27614.1;
SQ SEQUENCE 1070 AA; 114180 MW; 75254DDDD5643A5 CRC64;

Query Match 27.0%; Score 1823.5; DB 5; Length 1070;
Best local similarity 32.4%; Pred. No. 5.5e-139;
Matches 362; Conservative 155; Mismatches 416; Indels 183; Gaps 34;

Oy 21 GTASPLNEDNVNCSHMSYSTVQESYPHFEDQYYT-----SCTDILNMFECTRHR 73
Db 35 GTEP-----QSHVCT-----VKTIVDY--ELKVIHYVYVNDPQLNPLRGLFGC----- 80
Oy 74 VSYRTAYRHGKTYRRK-----SQCCEFTSEGE-MCVPHCADKCVHGRCTAPYTC 124
Db 81 -----TVERGQKASQYORLVRKKEKVKCCQBYOYTKHFCILPDCNPPCKGKCKIEPKC 136
Oy 125 QCEPGMGTCNSSACDGDHMGPHCTSRCCCKNGALCNPLTGACHCAAGFRGMRCEDECO 184
Db 137 EDDPGYGGKYCASSCSVGTWGLGSKSCDCENGANCDBELGCTCTSGFQGEKCEKPCPD 196
Oy 185 GTYNDCHQRCQCONGATCDHYTGECRCRPGYTGAFCEDLCPGKHGPECQRCPCQNGG 244
Db 197 NKMGPNVCYKSCPCQNGKCNK-EGKVCVSDGGEFCLNKEBEKFGABCKFECCQNGA 255
Oy 245 VCHAYTGECSCPSGMGTVCQPCPEGRFGKNSQEQCHNGSTCDAAATGCHQCSFGYTG 304
Db 256 TCDNTNGKCIKSGHGLACENESVGFSGCTCKCCLLNQNDSSSGCKCI-GWTG 314
Oy 305 ERCDCECPVGTGYLCAETCCV-----NGKCYHVSAGACICEAGFAGEBCEARLCPGEG 358
Db 315 KHCIDGCGRGRGLQCKNCTCPLGLEFSDNASCDARTGQOCQCEGYGPRKDEKCAE 374
Oy 359 LYGIKDCRCPCHLENTSCHPMGEGCACRPMGLYCNETCSPGTYEAC--QOICSQ 416
Db 375 QYGADESCYKTCYVRENTLMACPNTGFCRCRKGFGYDNCLELACKSKSYPNCEKQMDMN 434
Oy 417 NGADCDSTGTGCTCAPRGKIDCSNRCPLGTYGTCNCSRCCKKNAY--CSPPDSCCTCA 475
Db 435 HASECNPEGTGSCVCPGRTGKNCSEPCPLDTPGPAHQCCQNOGVCDDGADGCGQCDR 494
Oy 476 GWHGVDCSIRPSGTGFGCNLTCCOLNGACANTLDGTCTCAPGMRGECCELPCCDGTG 535
Db 495 GWTGHRCHHCRPADPFGANCKEKRCKRPGKIGCDPTTGECTCPAGIAGANCOIDGCESSYG 554
Oy 536 LNCAERCDSCSHADGCHPTTGHCRCPLGWSGVHCDSCVCAEGMRPNCSLPCYCKNGASCSP 595
Db 555 PGCKLHCKCVNGK-CDKETGECTCQPGFFGSDCSTYCSKGYGSECELSPCSD-ASCSK 612
Oy 596 DDGICECAPRGRTTCQATCSPGFYGHRCQOTCPCHVSSGPRCHH---TGLC-DCLGCF 651
Db 613 QTGKCLPLGTGKGVSCDDCKDPNTFGFLQETV-----TPSCASTDRKNGCCLSPGGS 667
Oy 652 TGALENEVCPSGRFGKNGAGICTCTNNGTCNPIDRSQCCYPMIGSDSCSPCPAHMPPN 711
Db 668 SGIRHEHNCRPAISYDGGQOVCSCADHGCBPTTGECLCEBCHYHKTCSSEKCPDCKYTG 727
Oy 712 CIHTC-NCNNAFCSAYDGECKCTPGWTGLYCTQRCPLGFTGKDCALICQCONG-ADCDH 769
Db 728 CALDPCKASGSTCDHNLICIPALGALCTRCBSAGFWNGCRQVCRCTSEVKQCA 787
Oy 770 ISGQCTTGFMGRHCEOKPSGTGYGCGROICDCLNNSCTCHITGYCSPGMKA 827
Db 788 QTGECSPAGFOGDRCDPCEDGYTGPPCIRKCKQGTATSSCNVSGACHCHPEFTGEF 847
Oy 828 CDOAGVILVNLNLSRST-----ALP-----ADSYQIGALAG----- 861
Db 848 C-----HALCPSTFGKCSKRCPRDGCSDGVECDAAIGCHVDMSCGKAKQE 896
Oy 862 -----IILVLVLF-----LIALFTYHKK--QKGESSMPAVTYTPAMRV 901
Db 897 FEALNGAGRSTGLTWFFVLLIIVALLCGGLTALF--YNNKYOKEXDPDMPTVSF----- 948

```

QY 902 VNADYITSGTLPHSNGNANSHFTNPSTHTLQCATSPHYNNRDMRYTKSKNNOLFVN 961
 Db 949 -----HKAPNDEGREFOPLX---SROSVFP---DSDAFSSENNNGH----- 986
 QY 962 LKAVNPGKRGVPDCTGTLPADMK-----HGGYLNELGAFGLDLSYMK-----SLK 1008
 Db 987 -----GGPPNGLLTLEEBELEENKRIHG-----RSMAGRGNNDYASLD 1023
 QY 1009 DLCKNSEYNSNCSLSSSENPYATIKDPYLIKPKSS 1044
 Db 1024 EVAGBESSSSASASARENPYADISSPDPYTONSA 1059
 RESULT 10
 ID 09XMD6 PRELIMINARY: PRT: 1111 AA.
 AC 09XMD6:
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Y47H9C.4 protein (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peleodierinae; Caenorhabditis.
 OC NCBI_Taxid=6239;
 RN [1]
 RA Sequence from N.A.
 RA Harris B.
 RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins J., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson R., Wellstock L., Wilkinson-Sproat J., Wohlman P.;
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163239;
 RA Zhou Z., Hartwig E., Horvitz H.R.;
 RT CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 RT Engulfment in C. elegans.";
 RL Cell 104:43-56(2001).
 DR EMBL: AL033657; CAA21739.1; -;
 DR EMBL: AF332568; AAG60061.1; -;
 DR HSSP: P05106; JUV2.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR002049; Laminin-EGF.
 DR Pfam: PR00008; EGF_9.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR SMART: SM00180; EGF_Lam; 6.
 DR SMART: SM00001; EGF_Like; 5.
 DR SMART: SM00261; FU; 2.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_15.
 DR PROSITE: PS01186; EGF_2; 11.
 KW EGF-like domain; Glycoprotein.
 SQ SEQUENCE 1111 AA; 118803 MW; A39F374C008E9874 CRC64;

QY 21 GTASPLNLEDPNVCSHMEISYTVQESYPAPEDOIYT-----SCTILNMFCKTRR 73
 Db 35 GTTEP---GGDHCT-----VKTIYVDY---ELKKVHTHYNDTECLNPLRFGOC----- 80
 QY 74 VSTRIVYRIGETMYRRK-----SQCCPGFESGE-MCVPHCKDKVHGHCIAIPNTC 124
 Db 81 ---TVEKRGOKKASYOQRLVKKREYVQCCDDGYQTKRHFCLEDCNPCKKGCIPGKC 136
 QY 125 QCEPBGWGTNCSACGDDHMGPHCTSCCOCKNLCALNPITGACHCAAGFGWGCEDRCG 184
 Db 137 ECDPFGTGGKTCASSGVGTWGLCCSKSCDENANCNDPELGTICTSGFQGERCEKPCPD 196
 QY 185 GTYNDCHQRCQCONGATDHYTEGRCRPRGYGAFCEBLCAPRKHPQCEORPCONG 244
 Db 197 NKWPGNCVASCPCQNGGKCK--EGKCVSGMGWGECECLNKEGRCGAGCKREFCNCQNGA 255
 QY 245 VCHHTVTEGSCSGSGMGWGYCGQPCPRGRGKNCSDQCCHNGSTCDAAATGQCHCSFGYTG 304
 Db 256 TCDNTNGKCIKCKSGYHGALCENECVGFQSGCTQKCDCLNMONCDSSECKCI--GMTG 314
 QY 305 EBCQDCPVGTVGLCAETQCCV-----NGKCYHVSAGLCEAGFAGEBRCARLCPSG 358
 Db 315 KNCDDICSKRRFGLQCKQNTCTPGLFESDSNASCAKTKGQCCESGYKGPCKDERKDAE 374
 QY 359 LYGKCDKRCPCPLENTHSCHPMSGECACKPGWSGLYCNETCSPGFYGEAC--QOICSCQ 416
 Db 375 QYGADCSKTCTCVRENTLMCAPNTGRCRCRPGYGDNCELAGSKDSYGRNCEQANCDMN 434
 QY 417 NCADCSVTGKCTCAPGFGIGDCTSPCLATYGINCSSRQCKANDAV--CSPVDGCTGCA 475
 Db 435 HASECNETGSCVCKPRGKNCSEPCPLDFYPNCAHOCQCNQRYGVGDGADGKCCOR 494
 QY 476 GHWGVNCSIRPCSGTWGFCGNLTCCCLNGACNTLGTCTCARGWGEKCELPCCGDTG 535
 Db 495 GMTGHRCEHHPADPTGANCERKCKPKGICGPIYBECTCPAGLAGANDIGCPGSGTG 554
 QY 536 LNCABKDCSHADGCHPTTGHCRLPGWSGVHDSVABGRMGPNCSLPCYKNGASCP 595
 Db 555 PGCKLHKCVKNGK--CKETGECTCGQFGSDCTSCSKGKYESCELSGPCSD--ASCK 612
 QY 596 DDCICCAQGRFRTTQORLCSPEFYGRHSQRCPOCVHSSGPCNHL---TGLC-DLBRF 651
 Db 613 QTKCKCLPGLTKVSDQCDPMTFGLQETV---PSPASDPNPNVGLVCLSPPS 667
 QY 652 TGAICNEVPSGRFGRKNCAGICTCTNNGTGCTNDRSCQCPGWSIGSDSOPCPAPMGPN 711
 Db 668 SGHCHENCPAGSYGDCQOVCSADGHGCDPTTGECICEBPGYHGTCKSEKCPDKRYTG 727
 QY 712 CIHTC--NCHNGAFCSAYDECKCTPGMTGLYCTQRCPLGFGKDCALICQONG--ADCDH 769
 Db 728 CALDCPKCASGSCDHIINGLICIPALLEGALCTRPCSAGFWGNGCQVRCSTSEYKQNA 787
 QY 770 ISGQCTCRGFMKRNHCEQCPSTYTYGCRQICDCLNNT--CDHTGTGYCSFGKAGAR 827
 Db 788 QTEGCSCPAGFQDRCDCPEDGYDPDCKCKCGYATSSCNRVSGACHPGFTGEF 847
 QY 828 CDQAGVLIYGNLSLSRTST-----ALP-----ADSYOIGATAG----- 861
 Db 848 C-----HALCPSTFGLKCKECPKDCGCGIECDALIGCHYDQMSCKRAOE 896
 QY 862 -----IITLVVLF---LALFITYRRK--QKRESSPAVYTPAMRY 901
 Db 897 FEALNGAGRSTGLTWFFVLLIYALCGGLIALF--YRNKYQKEKDPDPVTSF----- 948
 QY 902 VNADYITSGTLPHSNGNANSHFTNPSTHTLQCATSPHYNNRDMRYTKSKNNOLFVN 961
 Db 949 -----HKAPNDEGREFOPLX---SROSVFP---DSDAFSSENNNGH----- 986
 QY 962 LKAVNPGKRGVPDCTGTLPADMKHGYLNELGAFGLDSYV-----GSKLKDIAKN-- 1013
 Db 987 -----GGPPN--GLTLLEEBELEENKRIHGSAAGRNNDY 1019

Query Match 26.8%; Score 1805.5; DB 5; Length 1111;
 Best Local Similarity 31.8%; Pred. No. 17e-137;
 Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

Oy	305	ERODCEPVGVGLVLAETQACV-----NGKCYHVSAGCLCEAFAERCEATLCEPG	358
Db	315	KHCDIGSGRGRFGLQCAQNTCTPGLERFSDSNASCARKTGQCCQCESSQKPKPCDERKDAE	374
Oy	359	LVGIKCDKRCPCHELENTHSCPHMSGRCACKPGMSGLYCNETSPGFGYBAC--QQICSCQ	416
Db	375	QYGAADCSKTCVACVENTLMKAPMTGFCRCRKPFGYGNCELASKDSYGNPCEKQAMCDWN	434
Oy	417	NGACDSDVTGKCTCABCFKIDISTCQPLTGYINSSSGCKNDAY--CSPVDSCTCKA	475
Db	435	HASCCNBTGSCVCKPRTKNCSEPRDLFTYRPNCAHQCCQAKQKGVGDGADGKCCQDR	494
Oy	476	GMHWGDCSIRPCSGTWFGGNCILTCOCLNGACNTLIGTCTCAVGRGKECELPCCDGTG	535
Db	495	GMTGHRREHNCRAPDTFANGCEKRCRCKPGKIGCDPIRTGECTCPAGLOGANCDIGCEGSG	554
Oy	536	LNCAERDDCSHADGCHTTHGHCRLCRLQWMSVHNDSCVACGRGRPNCSLPCTYKKNASCP	595
Db	555	PGCLHCKVCYNGK--CDMETECTCQPRFFSDSTTSCGKKYGESELSLCPSPD--ASCSK	612
Oy	596	DDGICECAPGRGTTTCORISPGFYGHRCSQTCRPCVHSSGRCNHT---TGLC--DCLPGF	651
Db	613	QTHGCLDPLRGYKGVSCQCKDDPMTFGLCEIY----TPSCASIDPRKNGVCLCLCPGGS	667
Oy	652	TGALCNCEVSGRFGKAKACIGICTCTNNNGTCNPIDRSCQCTPGWISGDSQPCPPRAHWGN	711
Db	668	SGHCEHNRCPSAGSGDCQDQVCSADGHCDDPRTGECICEPFGYHKGTCSEKCPDCKYGY	727
Oy	712	CIHHC--NCHNGAFCSAVDGCCKCTPGMTGLCYTORCPLRGFKDALLIGCCONG--ADDDH	769
Db	728	CALDCPPKASGSTDHINGLICTPAGLEGLMCTRPCSAGRWGNGCHQVCRCTSEYKCNVA	787
Oy	770	ISGCTCRTEGFMHRCBOKCPSGTYGYGCRQICDCLNNT--CDHITGYCYSPGMKGAR	827
Db	788	QTEGSCPPAGCGRCRCKPDEGTYGDPDCKIKKCKCGTATSCNRRVSGACHHPFTGPF	847
Oy	828	CDQAGVYIVGNLSTST-----ALP---ADSYQIGANAG-----	861
Db	848	C-----HALCESTFGLKCSKECPKDCGGDGYEDDAIIGCHVNDOSCKAKOE	896
Oy	862	-----IIILVLYLF-----LALFIYRRK--OKGESSNPATYTPAMRKV	901
Db	897	FEALNGAGRSTGLWPFVLLIYVALLCGGLIALF--YRNKKYQEKEXDPDMPTVSF-----	948
Oy	902	VNADYITISGLPHNSNGNANASHFTNPSTYITLQCATSPHYVNNRDMATYTSKKNQOLYV	961
Db	949	-----HAPNNDEGRFQNPFLY---SRQSVFP---DSQAFSENNGNHQ---	986
Oy	962	LKNVNPCKRGVPGDCTTLPADWKHGGYLNELGAFILDRSYMGKSLDKGKXSEVSSNC	1021
Db	987	-----GGPRP--GLLTLLEEEL-----ENKKIHGRSA	1011
Oy	1022	SLSSENPATIKDPVLLIKSS	1044
Db	1012	AGRGKNKPYADISSPDEPTONSA	1034
RESULT 13			
063404			
AC	063404	PRELIMINARY; . PRT; 220 AA.	
DT	01-NOV-1996	(Tremblrel. 01, Created)	
DT	01-NOV-1996	(Tremblrel. 01, Last sequence update)	
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)	
DE	(clone REMA) ORF (Fragment).		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
OX	NCBI_TaxId=101116;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=HOLZMAN; TISSUE=BRAIN;		
TX	MEDLINE=96235155; PubMed=8642059;		

Query	Subject	Score	Length	Positives	Indels	Gaps
Q1	200 GATCDHVTGECRCPPGYTGAFCEBDCPPKXHGQCEBQRCPCQNGVGHNTGECSCPSGM	19.4%	1309	DB 11	220	
Q2	2 GATCDHVTGECRCPSGYTGAFCEBDCPPKXHGQCEBQRCPCQNGVGHNTGECSCPSGM	19.4%	1309	DB 11	220	
Q3	260 MGTCVGGPCPBGRRGRKNCSDCCCHNGTGDATATGCGHSPGYTGERCODECPVGYGV	19.4%	1309	DB 11	220	
Q4	62 MGTCVGGPCPBGRRGRKNCSDCCCHNGTGDATATGCGHSPGYTGERCODECPVGYGV	19.4%	1309	DB 11	220	
Q5	320 CAETCCVNGGCKHCVHGACLCCEAGFAGERCEARLCPBGYIGTKCRKPCHELNTHSCH	19.4%	1309	DB 11	220	
Q6	122 CAETCCVNGGCKHCVHGACLCCEAGFAGERCEARLCPBGYIGTKCRKPCHELNTHSCH	19.4%	1309	DB 11	220	
Q7	380 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q8	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q9	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q10	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q11	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q12	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q13	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q14	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q15	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q16	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q17	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q18	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q19	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q20	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q21	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q22	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q23	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q24	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q25	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q26	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q27	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q28	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q29	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q30	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q31	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q32	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q33	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q34	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q35	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q36	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q37	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q38	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q39	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q40	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q41	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q42	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q43	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q44	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q45	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q46	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q47	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	
Q48	182 PMSEGCACRPMGSLYCNCFSCPGFTGACQOCLSCQNG	19.4%	1309	DB 11	220	

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart J.H., Guan P., Harris M.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
 RA Jaisli M., Kalush F., Kaipen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervulov G., Mlshina N.V., Modyaty C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relneert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spredling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weltschbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003472; AAF47553.1; -.
 DR HSSP; P02468; IKLO.
 DR FLYbase; FBgn0035261; CG18172.
 DR InterPro; IPR001774; DSL.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF_8.
 DR PRINTS; PRO0011; EGF_LAMININ.
 DR SMART; SM00051; DSL; 1.
 DR SMART; SM00180; EGF_Lam; 3.
 DR SMART; SM00001; EGF_Like; 6.
 DR PROSITE; PS00022; EGF_1; 7.
 DR PROSITE; PS01186; EGF_2; 9.
 DR EGF-like domain; Glycoprotein.
 SO SEQUENCE 434 AA; 46417 MW; 26B5FD1E5C7970E4 CRC64;

Query Match 18.6%; Score 1252; DB 5; Length 434;
 Best Local Similarity 45.6%; Pred. No. 3, 6e-93;
 Matches 190; Conservative 55; Mismatches 170; Indels 2; Gaps 1;

OY 151 RCGCKGALCNPTGACGCAAGRGRCEDRCQGTGYGNDGRCQCCQNGATCDHTGRC 210
 Db 2 QCCCLNNAVCEPSPSGDCECAKGTGACADICPEGFFGANCSEKRCENKCHHVSCEC 61
 OY 211 RCPGTYGAFCEDLCPGKHGRCPCQNGVCHYTGECSPSGMWTGCGPCE 270
 Db 62 QCAFGTGLCDMRCPDGRHGAQCQDDPCQNDGKQCPETGACMCNPGWTGVCANPCV 121
 OY 271 GRRGKNCQBCQCHNGCTDAGTGCCHSGPYTGERGRCQDECPVGTGYVLAETGCVNG 330
 Db 122 GSYGSCQSCCECYKCAPRHHTGCECPYGRGRCPECOUETGFCNCSMCDCAANDA 181
 OY 331 KCYHSGACICEAGFAGEGCEARLPEGLYGIKCDKRCPCHEHNTSHCPMSECAKCR 390
 Db 182 MCDRANGTCLCNPGWTGAKCAERICANKTGLDCNRTCEBDMHTDLCHETGNCQCSIG 241
 OY 391 WSLGVCNCTSPGTYGEACQOLSCQNGADCVTGKTCYAPGFKIDESTPCPLTGYI 450
 Db 242 WSSAQCTRPCTFLRYGPNCLELNCNKAQKSPVNGTCLCAPWRBPTCESESEPFTFG 301
 OY 451 NGSRRGCKNDAYCSFYDSSCTCKAGWHGVDSIRCPSGTWGCGNLTCQCLNGACNLT 510
 Db 302 DCALRDCQNGACKEPETSQCLCTAGMKNIKCDRPPDLNHFQGDCKAYVDCNNANACNP 361
 OY 511 DGTCTCAPGWRGCKELPCODGTGYLCAERQDC--SHADGCHPTGHGCKLPGWGS 565
 Db 362 NGSCCTAAGWTGRCERKCDTGKFGHCAKQCCQCDNNSLACATNGRCVCKODMGG 418

RESULT 15
 ID 013149 PRELIMINARY; PRT; 2447 AA.
 AC 013149;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Notch 2 (Fragment).
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nakamura T., Trowsdale J.,
 RT "Structure of Fugu Notch gene."
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB004829; BAA20535.1; -.
 DR HSSP; P00740; 1EDM.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF-II.
 DR InterPro; IPR000083; Fibinectn.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PRO0010; EGFBLD.
 DR PRINTS; PRO0011; EGF_LAMININ.
 DR PRINTS; PRO0012; ENTYPER.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 20.
 DR SMART; SM00001; EGF_Like; 14.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 8.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_35.
 DR PROSITE; PS01186; EGF_2; 29.
 DR PROSITE; PS01187; EGF_CA; 20.
 DR ANK repeat; Calcium-binding; EGF-like domain; Glycoprotein;
 KW Hydroxylation; Repeat.
 FT NON_TER
 SO SEQUENCE 2447 AA; 262541 MW; 1A8E2A372A085D84 CRC64;

Query Match 15.2%; Score 1024.5; DB 13; Length 2447;
 Best Local Similarity 26.3%; Pred. No. 8, 4e-74;
 Matches 320; Conservative 101; Mismatches 325; Indels 471; Gaps 70;

OY 6 NSCLSFICLLCWMIGTASPLNEDPNVCSHMSYSVVOESYP--HPDQIYYTCTDI 63
 Db 99 DCLSSPCAN--GGTCTSLT-----GSLTSCSLPGTGRHCLNDTDECATRS 145
 OY 64 L--NPFKCTRHRYSRTAYRHGERTMYRRKSQCCPFEYSEGENC---VPHCADKCVHR 117
 Db 146 ICGNEGTCTNRTSGY-----KMCALGF--TKHCESSYIPCSPSPLNG 189
 OY 118 CLAPN-----TCCBERGKGTNCSSACDDGHHKPHCTSRQCKNGALCNPTGACGAAG 172
 Db 190 TCNONSSTYSCHCLPGLFGNTENNID--DCPHQCANRGTCTIDG---NTYNCQCPPE 244
 OY 173 FRGMRCEGCGTGYNDCH-QRCQCGATCDHTGGE--CRCPGYTGAFCEP----- 223
 Db 245 WTGQCHTEBY-----NECHLPNTCONGSTGSLNGSYCYVKNWSGLDCSENIDDD 298
 OY 224 -----LCPPGKHGPOC--EORC--PCQNGVVC--HHYTG--ECSC 255

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 04:45:58 ; Search time 8751 Seconds

(without alignments)
11383.717 Million cell updates/sec

Title: US-10-092-390-1

Perfect score: 3423

Sequence: 1 atggtattctcttggaacac.....gcagcagcagcagtgatga 3423

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:*

- 1: gb_da:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rtd:*
- 36: em_hlg_mam:*
- 37: em_hlg_vit:*
- 38: em_sy:*
- 39: em_hlgo_hum:*
- 40: em_hlgo_mus:*
- 41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3421.4	100.0	7522	AB058676	AB058676 Homo sapi
2	1692.4	49.4	2267	BC020198	BC020198 Homo sapi
3	1200.6	35.1	5702	AB058677	AB058677 Homo sapi
4	742	21.7	3281	HSMB805375	AL834326 Homo sapi
5	649.4	19.0	4290	AF444274	AF444274 Mus muscu
6	648.6	18.9	4539	AF41685	AF41685 Mus muscu
7	646.2	18.9	4482	AF440279	AF440279 Mus muscu
8	548.2	16.0	660	RATGFRD	L41686 Rattus norv
9	534.8	15.6	632	AK079681	AK079681 Sequence
10	456.8	13.3	4470	AK074121	AK074121 Homo sapi
11	400.8	11.7	5523	AB011532	AB011532 Rattus no
12	354.8	10.4	4501	AB011539	AB011539 Homo sapi
13	250.8	7.3	146206	AC026800	AC026800 Homo sapi
14	250.8	7.3	152765	AC008566	AC008566 Homo sapi
15	250.8	7.3	175144	AC010415	AC010415 Homo sapi
16	250.8	7.3	217221	AC008682	AC008682 Homo sapi
17	243	7.1	48202	AC012911	AC012911 Drosophi
18	243	7.1	172175	AC010038	AC010038 Drosophi
19	243	7.1	177583	AC105264	AC105264 Drosophi
20	243	7.1	257867	AC005557	AC005557 Drosophi
21	243	7.1	303191	AE003472	AE003472 Drosophi
22	241.8	7.1	192282	AC010424	AC010424 Homo sapi
23	239.8	7.0	2054	AK098809	AK098809 Homo sapi
24	216	6.3	3764	AF332568	AF332568 Caenorhab
25	197.2	5.8	163437	AC102173	AC102173 Mus muscu
26	192.4	5.6	153960	AC095558	AC095558 Rattus no
27	192.4	5.6	178298	AC094759	AC094759 Rattus no
28	180	5.3	62877	AC104414	AC104414 Mus muscu
29	166	4.8	178298	AC094759	AC094759 Rattus no
30	153	4.5	75342	CEY4799C	AC032657 Caenorhab
31	149	4.4	73791	AC091031	AC091031 Homo sapi
32	149	4.4	169638	AC009433	AC009433 Homo sapi
33	149	4.4	185854	AC027736	AC027736 Homo sapi
34	147.4	4.3	178862	AC011847	AC011847 Homo sapi
35	145.2	4.2	153960	AC095558	AC095558 Rattus no
36	140.2	4.1	177355	AC106503	AC106503 Rattus no
37	139.4	4.1	230668	AC121921	AC121921 Mus muscu
38	139.4	4.1	233204	AC092498	AC092498 Mus muscu
39	130.6	3.8	405	G51780	G51780 SHGC-84555
40	119.6	3.5	87428	CEY64G10A	AL110498 Caenorhab
41	119.6	3.5	297900	AC006704	AC006704 Caenorhab
42	118.8	3.5	73791	AC091031	AC091031 Homo sapi
43	118.2	3.5	3210	AF145617	AF145617 Drosophi
44	117.8	3.4	152765	AC008566	AC008566 Homo sapi
45	117.8	3.4	155630	AC015721	AC015721 Homo sapi

ALIGNMENTS

RESULT 1
AB058676
LOCUS
DEFINITION Homo sapiens mRNA for MEGF10 protein (R1MA1780), complete cds.
ACCESSION AB058676
VERSION AB058676.1 GI:14017776
KEYWORDS
SOURCE
ORGANISM Homo sapiens adult, hippocampus cDNA to mRNA, clone:pf01012.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (sites)
Nagase,T., Nakayama,M., Nakajima,D., Kikuno,R. and Ohara,O.
Prediction of the coding sequences of unidentified human genes. XX.
The complete sequences of 100 new cDNA clones from brain which code

for large proteins in vitro
JOURNAL DNA Res. 8 (2), 85-95 (2001)
MEDLINE 21245130
2 (bases 1 to 7522)
AUTHORS Nakayama,M., Nagase,T., Nakajima,D., Kikuno,R. and Ohara,O.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-2001) Manabu Nakayama, Kazusa DNA Research
Institute, Department of Human Gene Research; 1532-3, Yama,
Kisarazu, Chiba 292-0812, Japan (E-mail:manabu@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3915,
Fax:81-438-52-3914)

FEATURES

source
1..7522
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="pf01012"
/tissue_type="hippocampus"
/dev_stage="adult"
/note="vector:pBluescript II SK plus"
1..7522
/gene="MEGF10"
204..3626
/gene="MEGF10"
/note="KIAA1780 protein
gene encoding protein with multiple EGF-like-domains"
/codon_start=1
/product="MEGF10 protein (KIAA1780)"
/protein_id="BAB47409.1"
/db_xref="gi:14017777"
/translation="MVISNLSLFTICLLIHWIGTASPLNEDPNQSHMSYEV
DESYPHPDQIYYTCTDIPLNFKCTRIHRSYRTIYRIGETWYTRKSCQCKYALG
EMCVPHCKADCKYHRCIAPNTOQCEPFGMGWSSACGDDHDPICSTRQCKNKALE
NPITGACHCAAGFRGRCEDRCEDQYNGDCHQRCOCONGAICDVTBECRCPGYTG
AFCEIDLPKRGPOCEORPCONGVCHHTVDECSGNGMAYVCAQPCPFGRGKN
CSQECQNHNGTCDADATGCHSPGTGRCODEPVGIVGLCAETOCYVGGKCYH
VSGACIENGFAGGERGFAIRLCEPGLYGIKCDKRCCHLEHTHSPMGSEACRKGMS
GLYCMTSPFGYBACQOICSCONGACDSYTGCTCAPEGKIGDCSTPCPLGTIGI
NCSSRGCKNDVCSFVDSCTCKAGMHGVDSILPCSGTWFGNLTCQCLNGACN
TLDGTCTAPGWRGECBLPCQDGYGLNCAERCCSHADGHPPTGHCRLPGMSV
HCDYVCAGRMGPNCLSPCYCKNGASCSDDGICBAGFRGTCTORICSPFGYHRC
SOTCPQCVHSSGPNHITGLCPLPFTGALCNVCPGFRGKNCAGICTCTNNCTN
PDRSCOCYPGIGSDSOPCPAHMGNCIHTCNGHAGFCSAVDGCKTGMGTGL
YCTORCPGFYKDCALICCONGADCDHISGOCCTRGFPMRDEQCPSTYKXC
ROIICCLNNSTCDHTTCTCYGSPKHKGRCDAGIYIYONLSLSTRTALPADSTQI
GAIAGIILVLEVLFLFTIYRHQKQKESMSAVVITPAMRVNADYITISGLPH
SNGNANSHYFTNPSTHTLTOCATSPHVNNDRMVYTSKNQLEVNKNVNGRGP
VGDCTGLPADMKHGGYLNELGAFGLDRSYMGSKLDLKNSEYSSNCSLSSSENPY
ATIKDPVLIPEKSECGYEMKSPARDSPYAEIINNSTSANNVYEVEPTVVGVF
SNMRLSDPPVDLPKNSHPCHYDLIPVDRSSSPKODSGSSSSSSSSSP"

BASE COUNT

2130 a 1554 c 1697 g 2141 t

ORIGIN

Query Match 100.0%; Score 3421.4; DB 9; Length 7522;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 3422; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGTATTCTTTGAACTCATGCGCTGACCTTATTGTTTATGTTATGCGCACTGGATT 60
DB 204 ATGGTATTCTTTGAACTCATGCGCTGACCTTATTGTTTATGTTATGCGCACTGGATT 263
QY 61 GGGAGACATCACTGATCTTGAAGACCCCTAATGTGTGATGGACAGGGGAAACTAC 120
DB 264 GGGAGACATCACTGATCTTGAAGACCCCTAATGTGTGATGGACAGGGGAAACTAC 323
QY 121 TCAGTACTGTGACAGAGTCATACCCATCCCTTGTATCAATTTACTACAGAGCTGC 180
DB 324 TCAGTACTGTGACAGAGTCATACCCATCCCTTGTATCAATTTACTACAGAGCTGC 383
QY 181 ACTGACATTTAACTGTTTAAATGACAGCGGACAGAGTCAGTATGGAGAGCCTAT 240
DB 384 ACTGACATTTAACTGTTTAAATGACAGCGGACAGAGTCAGTATGGAGAGCCTAT 443
QY 241 CGACATGGGAGAAAGCTATGATAGGCGCAAGTCTCATGTTGTTCTCGATTTTATGAA 300

DB 444 CGACATGGGAGAAAGCTATGATAGGCGCAAGTCTCATGTTGTTCTCGATTTTATGAA 503
QY 301 AGCGGGGAAATGTGTCTCCCACTGCTGATTAATGTGTCCATGTGTGGCTGATGTCT 360
DB 504 AGCGGGGAAATGTGTCTCCCACTGCTGATTAATGTGTCCATGTGTGGCTGATGTCT 563
QY 361 CCAAAACCTGTACGTGTAGAGCTGTGGGAGGAGCAACATGTGCTCAGTGCCTGCAT 420
DB 564 CCAAAACCTGTACGTGTAGAGCTGTGGGAGGAGCAACATGTGCTCAGTGCCTGCAT 623
QY 421 GGTGATCACTGGGGTCCCACTGACACACCGGTGCTGCAAAAATGGGCTCTGTGC 480
DB 624 GGTGATCACTGGGGTCCCACTGACACACCGGTGCTGCAAAAATGGGCTCTGTGC 683
QY 481 AACCCCATCACCGGGGCTTGCACATGTGTGGGGGCTTCCGGGGGCTGGCGAGGAC 540
DB 684 AACCCCATCACCGGGGCTTGCACATGTGTGGGGGCTTCCGGGGGCTGGCGAGGAC 743
QY 541 CGCTGTAGCAGGAGGACCTATGTGTAACGACTGTCATCAGATGCGCAGTCCAGATGGA 600
DB 744 CGCTGTAGCAGGAGGACCTATGTGTAACGACTGTCATCAGATGCGCAGTCCAGATGGA 803
QY 601 GCCACCTGCGACACGTCACAGGGGGAATGCCGTGCCCAACAGATACACCGAGCTTC 660
DB 804 GCCACCTGCGACACGTCACAGGGGGAATGCCGTGCCCAACAGATACACCGAGCTTC 863
QY 661 TGTGAGATCTTTGTCTCTCTGTTAAACATGTGTCACAGATGAGAGAGATGCCCTGT 720
DB 864 TGTGAGATCTTTGTCTCTCTGTTAAACATGTGTCACAGATGAGAGAGATGCCCTGT 923
QY 721 CAAATGAGAGAGTGTGTCATACGTCACGTGAGAAATGCTCTTGCCTTCTGCTGTGATG 780
DB 924 CAAATGAGAGAGTGTGTCATACGTCACGTGAGAAATGCTCTTGCCTTCTGCTGTGATG 983
QY 781 GGCACAGTGTGTGTACGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGC 840
DB 984 GGCACAGTGTGTGTACGCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGCCTTGC 1043
QY 841 TCCCACTGCAATTAATGAGAGAGGAGTGTGATGTCGCACAGGCAATGTATTCAGTCA 900
DB 1044 TCCCACTGCAATTAATGAGAGAGGAGTGTGATGTCGCACAGGCAATGTATTCAGTCA 1103
QY 901 GGCATACAGAGGAGAGGTCGACAGATGATGTCCTGTTGGACCTATGGCTTCTGT 960
DB 1104 GGCATACAGAGGAGAGGTCGACAGATGATGTCCTGTTGGACCTATGGCTTCTGT 1163
QY 961 GGTGAGACCTGCGACGTGTGTCACAGGAGGAGAGTGTACAGGTGAGGGGCGATGCC 1020
DB 1164 GGTGAGACCTGCGACGT 1223
QY 1021 TGTGAAGCAGGCTTGTGCTGCGAGCGCTGCGAAGCAGCGCTGTCTGAGGGGCTCTAC 1080
DB 1224 TGTGAAGCAGGCTTGTGCTGCGAGCGCTGCGAAGCAGCGCTGTCTGAGGGGCTCTAC 1283
QY 1081 GGCATCAATGTGACAAACGCTGTCTCTCCACTTGGAAAACATCATATAGCTTACCCC 1140
DB 1284 GGCATCAATGTGACAAACGCTGTCTCTCCACTTGGAAAACATCATATAGCTTACCCC 1343
QY 1141 ATGTCTGAGAGTGTGCTGCGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1344 ATGTCTGAGAGTGTGCTGCGAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1403
QY 1201 TCTCTGATTTCTACGCGGAGACCTTGCACGAGATCTGACGTGCCAAAATGGGCGAGC 1260
DB 1404 TCTCTGATTTCTACGCGGAGACCTTGCACGAGATCTGACGTGCCAAAATGGGCGAGC 1463
QY 1261 TGTGACAGTGTACAGTGAAGTGTACCTGTGCCCGCAGATTTAAAGAAATTTAGCTCT 1320
DB 1464 TGTGACAGTGTACAGTGAAGTGTACCTGTGCCCGCAGATTTAAAGAAATTTAGCTCT 1523
QY 1321 ACCCATGCGCTCTGGGAACCTATGGGATTAACCTGTCTCTGCTGTGGCTGTAAAT 1380

```
Dh 1524 ACCCATCCCTGTCGGGAACCATATGGATTAACCTGCTCGCTGTCGGCTGTAATAAT 1583
Qy 1381 GATGACATCTGCTCTCTCTGTGGACGGGTCTTGTACTTGCAAGGACGCTGGCAGGGGTG 1440
Dh 1384 GATGACATCTGCTCTCTCTGTGGACGGGTCTTGTACTTGCAAGGACGCTGGCAGGGGTG 1643
Qy 1441 GACTGCTCAGATGATGTCAGATGGGACATGGGGCTTGGCTGTAACTTAACATGCCAG 1500
Dh 1644 GACTGCTCAGATGATGTCAGATGGGACATGGGGCTTGGCTGTAACTTAACATGCCAG 1703
Qy 1501 TGCCTCAACGGGGGAGCCTGCAACACCTGAGACGGGACCTGACGCTGTGCACTTGATG 1560
Dh 1704 TGCCTCAACGGGGGAGCCTGCAACACCTGAGACGGGACCTGACGCTGTGCACTTGATG 1763
Qy 1561 CGCGGGGGAATATGGAACTTCCCTGCGCAGATGGGACGTCACGCTGACCTGCTGAG 1620
Dh 1764 CGCGGGGGAATATGGAACTTCCCTGCGCAGATGGGACGTCACGCTGACCTGCTGAG 1823
Qy 1621 CGCTGCGACTGACGACGACGATGATGCTGCCACCTTACAGGGGCTATTGGCCGTCGCTC 1680
Dh 1824 CGCTGCGACTGACGACGACGATGATGCTGCCACCTTACAGGGGCTATTGGCCGTCGCTC 1883
Qy 1681 CGCGGATGTCAGGTGTCACATGTCAGCGTGTGTGCTGAGGAGACGCTGGGGCCCAAC 1740
Dh 1884 CGCGGATGTCAGGTGTCACATGTCAGCGTGTGTGCTGAGGAGACGCTGGGGCCCAAC 1943
Qy 1741 TGTCTCCCTGCTGCTACTGTAAAAATGGGCTTCATCTCTCCCTGATGATGGCATCTGC 1800
Dh 1944 TGTCTCCCTGCTGCTACTGTAAAAATGGGCTTCATCTCTCCCTGATGATGGCATCTGC 2003
Qy 1801 GAGTGTGACACGAGCTTCCGAGGACCACTTGTGACAGATGCTGCTCCCTGCTGTTTAT 1860
Dh 2004 GAGTGTGACACGAGCTTCCGAGGACCACTTGTGACAGATGCTGCTCCCTGCTGTTTAT 2063
Qy 1861 GGGCATGCTGACGACGACATGCCCACAGTGCCTTACACGACGAGGCGCCCTGCCACAC 1920
Dh 2064 GGGCATGCTGACGACGACATGCCCACAGTGCCTTACACGACGAGGCGCCCTGCCACAC 2123
Qy 1921 ATCACCGGCTGTGATGCTGTGCTGGCTTACAGGCGGCTCTGTGAATGATGCT 1980
Dh 2124 ATCACCGGCTGTGATGCTGTGCTGGCTTACAGGCGGCTCTGTGAATGATGCT 2183
Qy 1981 CCCAGTGGCAGATTTGGAAAACTGTGACAGAAATTTGTACTGACCAACAGCAAGCAAC 2040
Dh 2184 CCCAGTGGCAGATTTGGAAAACTGTGACAGAAATTTGTACTGACCAACAGCAAGCAAC 2243
Qy 2041 TGTAACTCCATTTGACATCTTTGTGACGTGTAACCCGGTGGATTGGCATGACTGCTCT 2100
Dh 2244 TGTAACTCCATTTGACATCTTTGTGACGTGTAACCCGGTGGATTGGCATGACTGCTCT 2303
Qy 2101 CAACGATGTCACGCTGCCCATGCGGAGCCCAAACTGCATCCACAGTGCATGCTGCATAT 2160
Dh 2304 CAACGATGTCACGCTGCCCATGCGGAGCCCAAACTGCATCCACAGTGCATGCTGCATAT 2363
Qy 2161 GGAGCTTCTGACGCGCTTACGATGGGAAATGTAATGCACTCTGCTGACAGAGGCTC 2220
Dh 2364 GGAGCTTCTGACGCGCTTACGATGGGAAATGTAATGCACTCTGCTGACAGAGGCTC 2423
Qy 2221 TACTGCACTGACAGATGTCCTCTAGGGTTTATGAAAAAGATTGTGCACTGATATGCCAA 2280
Dh 2424 TACTGCACTGACAGATGTCCTCTAGGGTTTATGAAAAAGATTGTGCACTGATATGCCAA 2483
Qy 2281 TGTCAAAAAGAGGCTGATGACACATTTCTGGGACGATGCTGCGGACGCTGATGTC 2340
Dh 2484 TGTCAAAAAGAGGCTGATGACACATTTCTGGGACGATGCTGCGGACGCTGATGTC 2543
Qy 2341 ATGGAGCGGCACTGTGACAGAAAGTGCCTTACAGAACATATGGCTATGGCTGTCAG 2400
Dh 2544 ATGGAGCGGCACTGTGACAGAAAGTGCCTTACAGAACATATGGCTATGGCTGTCAG 2603
Qy 2401 ATATGTGATGTGCTGAACAATCTCCACCTGCGACCAATCATGCGACTGTTACTGACG 2460
Dh 2604 ATATGTGATGTGCTGAACAATCTCCACCTGCGACCAATCATGCGACTGTTACTGACG 2663

Qy 2461 CCCGATGGAAGGAGCCGAGATGTATCAAGTGGTGTATATATGTTGAAATGTGAC 2520
Dh 2664 CCCGATGGAAGGAGCCGAGATGTATCAAGTGGTGTATATATGTTGAAATGTGAC 2723
Qy 2521 AGCTTAACCCGAAACAGTACTGCTCCCTGCTGATGTTCTTACATGAGGGGCTATGCA 2580
Dh 2724 AGCTTAACCCGAAACAGTACTGCTCCCTGCTGATGTTCTTACATGAGGGGCTATGCA 2783
Qy 2581 GGCATCATATCTTCTGCTGCTAGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2640
Dh 2784 GGCATCATATCTTCTGCTGCTAGTTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2843
Qy 2641 CACAAGCAGAAGGAGAAAGAAATCAAGCATGCGCAGAGTATCTTACACCCCTGCTATGAG 2700
Dh 2844 CACAAGCAGAAGGAGAAAGAAATCAAGCATGCGCAGAGTATCTTACACCCCTGCTATGAG 2903
Qy 2701 GTGCTCAATGCAAGATTAATCAATTTACAGTAACCTTCTCTCTCTCTCTCTCTCTCTCTCT 2760
Dh 2904 GTGCTCAATGCAAGATTAATCAATTTACAGTAACCTTCTCTCTCTCTCTCTCTCTCTCT 2963
Qy 2761 AATAGCACTACTTACCAATCCAGTTACCAACAGCTCACCCAGTGTGCCATCCCT 2820
Dh 2964 AATAGCACTACTTACCAATCCAGTTACCAACAGCTCACCCAGTGTGCCATCCCT 3023
Qy 2821 CACGTCAACACAGGAGACGATGACTGTGACAGAGTCAAAAAAACAATCACTGTTGTG 2880
Dh 3024 CACGTCAACACAGGAGACGATGACTGTGACAGAGTCAAAAAAACAATCACTGTTGTG 3083
Qy 2881 AATCTTAATAAATGTAACCCCTGAGAGAGAGGCTGAGGGGACTGCACTGGGACATTTG 2940
Dh 3084 AATCTTAATAAATGTAACCCCTGAGAGAGAGGCTGAGGGGACTGCACTGGGACATTTG 3143
Qy 2941 CGGGTGAACGAAACATGCGGGCTACTCTCAACGAGCTGCTGTTTGGACTTGACAGA 3000
Dh 3144 CGGGTGAACGAAACATGCGGGCTACTCTCAACGAGCTGCTGTTTGGACTTGACAGA 3203
Qy 3001 AGCTTATGGGAAATCCTTAAAGACCTGGGAAAGAAATTCGATTAATTCAGATTAAC 3060
Dh 3204 AGCTTATGGGAAATCCTTAAAGACCTGGGAAAGAAATTCGATTAATTCAGATTAAC 3263
Qy 3061 TGTCTCCCTTACGAGTTCGAGAACCCATATGCGCATTTAAAGACCCACTGTACTTATC 3120
Dh 3264 TGTCTCCCTTACGAGTTCGAGAACCCATATGCGCATTTAAAGACCCACTGTACTTATC 3323
Qy 3121 CCGAAACCTCAGAGTGTGTTATGTGAGATGAATGCGCGGACGAAAGATTCGCCA 3180
Dh 3324 CCGAAACCTCAGAGTGTGTTATGTGAGATGAATGCGCGGACGAAAGATTCGCCA 3383
Qy 3181 TATGAGAGATCAATTAATCAACTCAGCCAAACAGAAATGCTATGAATGAACCTTACA 3240
Dh 3384 TATGAGAGATCAATTAATCAACTCAGCCAAACAGAAATGCTATGAATGAACCTTACA 3443
Qy 3241 GTGAGTGTGTCAGAGAGTATTCAGCAATAATGCGCTCTCTCCAGATCCATATGAC 3300
Dh 3444 GTGAGTGTGTCAGAGAGTATTCAGCAATAATGCGCTCTCTCCAGATCCATATGAC 3503
Qy 3301 CTCCCAAGAACAGTCACATCCCTTGTATATGACCTGCTGCGCAGTCCGAGACAGTTCA 3360
Dh 3504 CTCCCAAGAACAGTCACATCCCTTGTATATGACCTGCTGCGCAGTCCGAGACAGTTCA 3563
Qy 3361 TCTTCCTCTAAGCAAGAGAGAGTGGAGTGGACACACACAGCAGCAGCAGAGTGA 3420
Dh 3564 TCTTCCTCTAAGCAAGAGAGAGTGGAGTGGAGTGGACACACACAGCAGCAGCAGTGA 3623
Qy 3421 TGA 3423
Dh 3624 TGA 3626

RESULT 2
BC020198
LOCUS BC020198 2267 bp mRNA linear PRI 22-JAN-2002
```

DEFINITION Homo sapiens, similar to MGF10 protein, clone MGC:31776
 IMAGE:4904255, mRNA, complete cds.
 ACCESSION BC020198
 VERSION BC020198.1 GI:18044365
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens.
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 2267)
 REFERENCE Strausberg, R.
 TITLE Direct Submission
 AUTHORS Submitted (19-DEC-2001) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA
 JOURNAL NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 REMARK Contact: MGC help desk
 COMMENT Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Genome Sequence Centre,
 BC Cancer Agency, Vancouver, BC, Canada
 Info@bgsc.bc.ca
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
 Letlicia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
 Nesh, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
 Schein, Duane Smilans, Michael Smith, Lorraine Speede, Jeff Stott,
 Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
 George Yang, Scott Zuyderduyn, Marco Marra.
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://Image.lnl.gov>
 Series: IRAL Plate: 40 Row: 0 Column: 4
 This clone was selected for full length sequencing because it
 passed the following selection criteria: matched mRNA gi: 14192942.
 FEATURES
 source
 Location/Qualifiers
 1..2267
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:31776 IMAGE:4904255"
 /tissue_type="Muscle, fibroblastsarcoma"
 /clone_11b="NH-MGC_17"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 280..1983
 /codon_start=1
 /product="Similar to MGF10 protein"
 /protein_id="AAH20198.1"
 /db_xref="gi:18044366"
 /translation="MVISLNSCLSFICLLCHMIGTASPLNLEDPNVCMSHESYVT
 OESYHPFDIYYSSTIDILNMFKTRHVSRTYRTHGKEMRYRKSCOCCEPESG
 EKVCHCAKCYHRCIAPNTCCCEPBGKGTSCSACGSDHGPCTSRCKKALC
 NPITACACAGFRGWRCEDEKQDQNDHQRCQCNAGATCDVTEBCRPAYTG
 AFCEDLCPGKAGHPQCEQRCPOQNGVYCHHYTGSCSPGNMGVTCGCPCEGRGKN
 CSOECCHNGGTCDNAEQCHCSPEYTERODECPVYGVLCAGETOCQVGVGCKYN
 VSGALCEAGFAGEGCEARLCEPLGYIKDKRCRCHLENTSHCHPMGSECKRGS
 GLYENETSPGFYGAACODICSCONGACDSDVCTCCTCAGPCKGIDCSPTCPLGYGI
 NCSSRCGCKNANVCSPYVDSCTCRKAGHGVDSICSPGTFMPCGNLTCQCLNGACN
 TLDFGCTCAPGMRGKCELPQDDGYGILNCAERDCSHADGCHPTTGHCRLPGWSGL
 F"
 BASE COUNT 543 a 548 c 631 g 545 t
 ORIGIN
 Query Match 49.4%; Score 1692.4; DB 9; Length 2267;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1693; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 0Y 1 AAGGATATTCTTGAACGATGCGACCTTATTTGTTATGTTATGACCACTGGATT 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 280 AAGGATATTCTTGAACGATGCGACCTTATTTGTTATGTTATGACCACTGGATT 339
 QY 61 GGGACAGCATCAACCTTGATCTTGAAGACCCATATGTGTGAGCCATGGGAAAGCTAC 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 340 GGGACAGCATCAACCTTGATCTTGAAGACCCATATGTGTGAGCCATGGGAAAGCTAC 399
 QY 121 TCAGTACTGTGCAGAGTCAATACCCACATCCCTTGTATCAATTTACTACAGAGCTGC 180
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 400 TCAGTACTGTGCAGAGTCAATACCCACATCCCTTGTATCAATTTACTACAGAGCTGC 459
 QY 181 ACTGATATTCTAAACCTGTTTAAATATCAGCGGCACAGAGTACGATATGGACAGCTAT 240
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 460 ACTGATATTCTAAACCTGTTTAAATATCAGCGGCACAGAGTACGATATGGACAGCTAT 519
 QY 241 CGACATGGGAGAGAGACTATGATATAGCCGCAAGTCAAGTGTCTCTGATTTATGAA 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 520 CGACATGGGAGAGAGACTATGATATAGCCGCAAGTCAAGTGTCTCTGATTTATGAA 579
 QY 301 AGCGGGGAAATGTGTCTCCCACTGTGATTAATGTGCCATGCTGTATGCT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 580 AGCGGGGAAATGTGTCTCCCACTGTGATTAATGTGCCATGCTGTATGCT 639
 QY 361 CCAACACCTGTCAGTGTAGCTGCTGGGAGGAGACCAACTGCTCAGTGCCTGAT 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 640 CCAACACCTGTCAGTGTAGCTGCTGGGAGGAGACCAACTGCTCAGTGCCTGAT 699
 QY 421 GGTGATCACTGGGCTCCCACTGACACACCGGCTGCAGTGCAGTAAATGGGCTCTGTGC 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 700 GGTGATCACTGGGCTCCCACTGACACACCGGCTGCAGTGCAGTAAATGGGCTCTGTGC 759
 QY 481 AACCCATACCGGGGCTTCCCACTGTGCGGGGCTTCCGGGGCTGGGCTGCGAGAC 540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 760 AACCCATACCGGGGCTTCCCACTGTGCGGGGCTTCCGGGGCTGGGCTGCGAGAC 819
 QY 541 CGCTGTGACAGGGACCTATGATGATACAGTGCATCAGAGATGCCAGATGGA 600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 820 CGCTGTGACAGGGACCTATGATGATACAGTGCATCAGAGATGCCAGATGGA 879
 QY 601 GCCACCTGGACACGTCACAGGGGGAATGCCGTGCCACACAGATACACCGAGCCTTC 660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 880 GCCACCTGGACACGTCACAGGGGGAATGCCGTGCCACACAGATACACCGAGCCTTC 939
 QY 661 TGTGAGATCTTGTCTCTCTGTAAACATGTGCCACATGTGAGAGAGATGCCCTTGT 720
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 940 TGTGAGATCTTGTCTCTCTGTAAACATGTGCCACATGTGAGAGAGATGCCCTTGT 999
 QY 721 CAAATGAGAGATGTGTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 780
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1000 CAAATGAGAGATGTGTATCATCATCATCATCATCATCATCATCATCATCATCATCAT 1059
 QY 781 GGCACAGTGTGTGTGACGCTTGCCTCCGAGAGGCTGCTTGGAAAGATGTTCCCAAGA 840
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1060 GGCACAGTGTGTGTGACGCTTGCCTCCGAGAGGCTGCTTGGAAAGATGTTCCCAAGA 1119
 QY 841 TGGCAATGGCATTAAATGAGAGAGAGCTGTATGCTGCCACAGCCAAATGTATTCAGTCCA 900
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1120 TGGCAATGGCATTAAATGAGAGAGAGCTGTATGCTGCCACAGCCAAATGTATTCAGTCCA 1179
 QY 901 GGATACACAGGGGAAAGGAGTGCAGATGAGTGTCTGTGGGACCTATGGCGTTCCTGT 960
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1180 GGATACACAGGGGAAAGGAGTGCAGATGAGTGTCTGTGGGACCTATGGCGTTCCTGT 1239
 QY 961 GGTGAGACCTGCGCAGTGTGTCAACGAGAGGAGTGTATTCACAGTGAAGCGGCGCATGCCCTC 1020
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1240 GGTGAGACCTGCGCAGTGTGTCAACGAGAGGAGTGTATTCACAGTGAAGCGGCGCATGCCCTC 1299
 QY 1021 TGTGAAGCAGGCTTGTCTGTGGCAGCGCTGCGAAGCAGCGCTGTCTTAAGGGGCTCTTAC 1080
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1300 TGTGAAGCAGGCTTGTCTGTGGCAGCGCTGCGAAGCAGCGCTGTCTTAAGGGGCTCTTAC 1359
 QY 1081 GGCATCAATGTGACAAAGGAGTGTCCCTGCGCACTTGGAAACACATCATATGCTGTACCGCC 1140
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1360 GGCATCAATGTGACAAAGGAGTGTCCCTGCGCACTTGGAAACACATCATATGCTGTACCGCC 1419

```

OY 1141 ATGCTGAGAGTGTGCTGCAAGCCGGGCTGGTCAAGACTTACTATATAGACATGT 1200
    |||||||
Db 1420 ATGTCTGGAAGTGTGCTTCAAGCCGGGCTGGTCAAGACTTACTATATAGACATGT 1479
OY 1201 TCTCTGAGATTCACGGGGAAGCTTGGCCACAGATCTGCAGTCCCAAAATGGGGCAGAC 1260
    |||||||
Db 1480 TCTCTGAGATTCACGGGGAAGCTTGGCCACAGATCTGCAGTCCCAAAATGGGGCAGAC 1339
OY 1261 TGTGACAGTGTACGTGAAAGTCCACCTGTGCCCAAGATTCAAAGAAATGACTGCTCT 1320
    |||||||
Db 1540 TGTGACAGTGTACGTGAAAGTCCACCTGTGCCCAAGATTCAAAGAAATGACTGCTCT 1599
OY 1321 ACCCATGCGCTCTGGGAACTATGGATTAACGTTCCTCTGCTGGCTGTAAATAAT 1380
    |||||||
Db 1600 ACCCATGCGCTCTGGGAACTATGGATTAACGTTCCTCTGCTGGCTGTAAATAAT 1659
OY 1381 GATGCAAGTGTCTCTCTGTGAGAGGGTCTTGTACTTGCACAGCAGGCTGGCAGGGGTG 1440
    |||||||
Db 1660 GATGCAAGTGTCTCTCTGTGAGAGGGTCTTGTACTTGCACAGCAGGCTGGCAGGGGTG 1719
OY 1441 GATGCTCCATCAGATGTCCCACTGTCACATGGGGCTTTGGCTGTAACTTAACATGCCAG 1500
    |||||||
Db 1720 GATGCTCCATCAGATGTCCCACTGTCACATGGGGCTTTGGCTGTAACTTAACATGCCAG 1779
OY 1501 TGCCTCAAGGGGGAGAGCTGCAACACCCCTGGAGGGAGCTGCAAGTGTGACCTGGATGG 1560
    |||||||
Db 1780 TGCCTCAAGGGGGAGAGCTGCAACACCCCTGGAGGGAGCTGCAAGTGTGACCTGGATGG 1839
OY 1561 CGCGGGGAGAATGCAAACTTCCCTGCGCAGAGTGGCAGCGTACGGGCTGAACTGTGTGAG 1620
    |||||||
Db 1840 CGCGGGGAGAATGCAAACTTCCCTGCGCAGAGTGGCAGCGTACGGGCTGAACTGTGTGAG 1899
OY 1621 CGCTGGACCTGACCGACAGATGGTGTGACCTTACCAAGCGGCTTCCCGCTGCTC 1680
    |||||||
Db 1900 CGCTGGACCTGACCGACAGATGGTGTGACCTTACCAAGCGGCTTCCCGCTGCTC 1959
OY 1681 CCGGATGTGTGAG 1694
    |||||||
Db 1960 CCGGATGTGTGAG 1973

```

RESULT 3
AB058677
LOCUS AB058677 5702 bp mRNA linear PRI 10-MAY-2001
DEFINITION Homo sapiens mRNA for MEGF11 protein (K1A1781), complete cds.
ACCESSION AB058677
VERSION AB058677.1 GI:14017778
KEYWORDS
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone:fg06971.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (sites)
 Nagase, T., Nakayama, M., Nakajima, D., Kikuno, R. and Ohara, O.
 Prediction of the coding sequences of unidentified human genes. XX.
 The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro
 DNA Res. 8 (2), 85-95 (2001)
JOURNAL 21245130
MEDLINE 2 (bases 1 to 5702)
REFERENCE Nakayama, M., Nagase, T., Nakajima, D., Kikuno, R. and Ohara, O.
AUTHORS Direct Submission
TITLE Submitted (27-MAR-2001) Manabu Nakayama, Kazusa DNA Research
JOURNAL Institute, Department of Human Gene Research; 1532-3, Yama,
 Kisarazu, Chiba 292-0812, Japan (E-mail:manabu@kazusa.or.jp,
 URL:http://www.kazusa.or.jp/huge, Tel:81-438-52-3915,
 Fax:81-438-52-3914)
FEATURES
 Location/Qualifiers
 1..5702
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="fg06971"

```

/tissue-type="brain"
/dev-stage="fetal"
/notes="vector: pBluescript II SK plus"
1..5702
/gene="MEGF11"
160..3069
/feature="MEGF11"
/feature="K1A1781 protein"
Start codon is not confirmed. fg06971 cDNA clone for
K1A1781 has a 119bp insertion after the position 2642 of
the sequence of K1A1781
gene encoding protein with multiple EGF-like domains"
/codon_start=1
/product="MEGF11 protein (K1A1781)"
/protein_id="BAB47410.1"
/db_xref="GI:14017779"
/translation="MHTPSIRITHTDAQTSSTGSSAPGALCTEBCVHRCVSPDTC
CEPVGSGPDCSSGSDSHWPGHSCNRCQONALPTIGACVCAAGFPMRCCEILCA
PGTHGKGOPLPCQHRGASCDPRAGKICAPATGYCELPSPSHGHELCRPOC
NGCTCHHTGEBACPRGWTGAYCAQPCPRGTCGONSCDPCPHHGGOCDHYTGCHT
ACTMGDRCEBEPFSGFQCSQRCDCHNGGCSPTTGACEPEPTKPRQDERLCPE
GLHGPCTLPCEPDADNTISCHPVTAAGTCQDQSGHGHNECSPVGYDGLQPCPC
ONGADCHSTTGCTCAGPRFMGEVCAVSCAAGTYGPNCSSICGNNGTCSIPVDSGCT
KEGMOGLDCTLPSPSTWGLNENSCCANAGTSPIDSCSTPGMLDTCLEPCD
GTFGLNCSHEDCSHADGCDPVTGHCCCAAGTGIRCDSTCPGRMGPGSYSCEN
GGSGPEDESCBAPGFERPLQRTCPPEFVGHGAGAPCPILCVHSSRPCHHSIIGEC
LPFGSALCNQYACAGTREGDQAOLCSANNCTSPIDSCCCTPGMIGKDCSONCP
GFWPACPRHACSHNGASCASADGACHCTPGWTGLPCTRCRAFFGKCKGVCCON
GASCDHISGKCTRTGFTGQHEORCAPETFGYGCQQLCEKNNSTCHVTCTVCSF
GFGKIGIDDAALAMELEINPYTKISPALGERSVAVAGIMLFFIVLVLLGFLFMR
RQKEKGRDLAPRVSTYTPAMRTSTDSYLSGACGDMDRONTYIMDKGFEDYKESYS
STGSLNSENSENYATIKDPIILTKLPSESYVEMKSPVMGSPYTDVPSLSTNNIT
EVEPIYSYVQBECCGHNSSYIQNAYLPRHSHTLPHHTDLLPVQSPANPSQDKOS"
BASE COUNT 1495 a 1500 c 1382 g 1325 t
ORIGIN
Query Match 35.1%; Score 1200.6; DB 9; Length 5702;
Best Local Similarity 69.2%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 739; Indels 6; Gaps 2;
OY 318 CCCCACCTGTGCTGATTAATGTGTCCATGCTGCTGATTCCTCAAAACCTGTGAGT 377
    |||||
Db 234 CGCCTGTGTACGAGAGAGTGTGACAGCCGCTGGCTTCCCGGACACTGTGCACTG 293
OY 378 TAGACCTGTGTGGGGAGAGACCAACTGCTTCACTGCTGCGATGATGATGAGGCTC 437
    |||||
Db 294 CGAGCTGTGTGGGGAGAGGCCCGGACGCTTCCACCGGCTGTGACAGGACGACACTGGGGCC 353
OY 438 CCACTGACACGCGGCTGCGCAGTGCAGAAATGGGGCTGTGCAACCCCATCCAGCGGGC 497
    |||||
Db 354 CCACTGACACGCGGCTGCGCAGTGCAGAAATGGGGCTGTGCAACCCCATCCAGCGGGC 413
OY 498 TTGCACTGTGTGCGGGCTTCCGGGGCTGCGCTCGAGAGACCGCTGTGAGCAGGGCAC 557
    |||||
Db 414 CTCGATGTGGCGCGCGGCTTCCGTGATGCGCTCGAGAGACTGTGCGCGCTGGCAC 473
OY 558 CTATGTGAAGCACTGTCATCAGAGATGCCAGTGTCCAGATGAGGCCACCTGCGACACAG 617
    |||||
Db 474 CCACGGCAAGGGATGTCAGCTGTGCGGCTGCGCAGTGCAGACAGGATGAGCTGCACCCCG 533
OY 618 CACGGGGAATGCGCGTGGCCACACAGATACACGAGAGCTTGTGAGAGATTTGTGCC 677
    |||||
Db 534 CGCGGCGAGATGCTTGTGCGACCTGTGACACCGCGCTTACTGCGAGAGAGCTGTGCC 593
OY 678 TCTGTGTAACAGTGTCAACAGTGTGAGCAGAGATGCCCTTGTGTAACATGAGAGAGTGTG 737
    |||||
Db 594 TCTGTGAGAGCAGTGTGACAGTGTGAGCTGCGCTCCCTGTGCAAAATGGGGCAGCTG 653
OY 738 TCAATCAGTGTGAGAGATGCTTGTGCGCTTGTGCTGATGAGAGGACAGTGTGTGTA 797
    |||||
Db 654 CCAACCATCATCTGCGGAGAGTGTGCTGCGCCCGGCTGAGAGGAGAGTGTGCTGCCA 713
OY 798 GCCTTCCCCCGAGGCTGCTTGGAAAGAACTTCCCAAGATGCACTGCAATTAATG 857

```


JOURNAL COMMENT

Submitted (09-JUL-2002) 1, D-85764 Neuberberg, GERMANY
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 Sequenced by DKFZ (German Cancer Research Center,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 This clone (DKFZp434L121) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heuberweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further
 information about the clone and the sequencing project is available
 at <http://mips.gsf.de/proj/cDNA/>.

FEATURES

source

1..3281
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp434L121"
 /tissue_type="testis"
 /clone_id="434 (synonym: htes3). Vector pSport1; host
 DH10B; sites NotI + SalI"
 /dev_stage="adult"
 1..3281
 /gene="DKFZp434L121"
 <1..1883
 /gene="DKFZp434L121"
 /note="Similarity to MEGF6 (Rattus norvegicus)"
 /codon_start=-3
 /product="hypothetical protein"
 /protein_id="CAD8994.1"
 /db_xref="GI:2173946"
 /translation="GRVPGAVAVAOQPCPPTFGONCSQDPCHHGSCDHTVAGCHCT
 AGYMGDRCEPEFSGFQSCORDCDHNGSCSPPTGACEPEGYGPGCOLPCE
 GLHSGCTLPCPDADNTISCHPYTGACTCOPKMGHNCSECPVYGGCOLPCE
 ONAGDSHTSGCTCARGMEVCAVCAAGTYGPNSSISCSNNGSTCPYVDSCTC
 KEHQGLDCTLPSPSTWGLNCSCTCANGAACSPIDGSCSPGLGDTCELPED
 GTFGLNCSHCDSHADGCDPVTHCCCLAGWGIKIDSDTCEPRMGNPNSVSCEN
 GSGSPEDSGCECAPGFRGPLCORICPPGYGHCAQPCPLCVSSSPHHISICPC
 LPFGSALCNQVAGGYGODCAOLCSCANNNGCSPIIDGSCQCPGWIGKDSACP
 GFNGPACFHACSHNGNSGAEADGACHCTGWTGLFCTORKPHILASQPLRIKCGIL
 ATGVYQTSREGGQNAAPGLVVPDSCTPRREHLCGSSRRDHIQIDKPRVLESGCK
 AAQOHFLGRIVGAIASVYKMAPAVTTSVASPAQAQSPGNTVSQVPEPLAMGVSTV
 SA"

CDS

gene

polya_site
 3232
 /gene="DKFZp434L121"
 BASE COUNT 749 a 967 c 886 g 679 t
 ORIGIN

Query Match

Best Local Similarity 69.2%; Pred. No. 1.8e-217;
 Matches 1012; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

QY 781 GGACAGTGTGTGTCACACCTTGCCTCCGAGAGGTGGCTTGGAAAGAACTGTCCCAAGAA 840
 DB 18 GGAGCAGTGTGTGCCACCTCGCCACACGAGGACATTGGCCAGAACTGACCGCAGAT 77
 QY 841 TGCAGTGCCTTAATGAGGAGCAGTGTGATCTGCACAGGCAATGTCTTGCAGTCA 900
 DB 78 TGTCTTGCACCAATGAGGAGCAGTGTGATCTGCACAGTGTGATCTGCAGTCA 137
 QY 901 GGATACACAGGGAACGCTGCAGAGATGATGTCTTGGAGCACTATGCGTCTCTGT 960
 DB 138 GGATACATGAGGAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 197
 QY 961 GCTGAGACCTGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 DB 198 TCACAGCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 257
 QY 1021 TGTGAGCAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 DB 258 TGTGAGCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 317
 QY 1081 GGATCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
 DB 318 GGCCAGGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 377

QY 1141 ATGTCGTGAGAGTGTGCTGCAACCGGGCTGTGTCAGACTCTACTGTATGATGATGAT 1200
 DB 378 GTAACTGAGTGTGTACTGTGCTGACACGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 437
 QY 1201 TCTCTGATATCTACGAGGAGTGTGCTGACAGAGATCTCAGACTCTGCAAAATGGGAGAC 1260
 DB 438 CTTGTGTGCTACTGTGAGGAGTGTGCTGACAGAGTGTGCTGACAGAGTGTGCTGACAGAG 497
 QY 1261 TGTGACAGTGTGATGAGGAGTGTGCTGACAGAGTGTGCTGACAGAGTGTGCTGACAG 1320
 DB 498 TGCCACAGCATCACTGGGGGCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 557
 QY 1321 ACCCATGCTCTGAGGAGTGTGCTGACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
 DB 558 GTTCTGTGTGAGCAGAGGAGTGTGCTGACAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 617
 QY 1381 GATGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440
 DB 618 GTTGTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 677
 QY 1441 GACTGCTCATCAATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 DB 678 GACTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 737
 QY 1501 TGCTCAACGGGGGAGCTGCAACACCTGAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1560
 DB 738 TGTGCAATGAGGAGTGTGCTGCAACACCTGAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTG 797
 QY 1561 CGCGGAGGAATATGCAACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
 DB 798 CTGGGAGGACACTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 857
 QY 1621 CGTGTGCACTGACAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
 DB 858 CACTGTGACTGACAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 917
 QY 1681 CGGGATGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
 DB 918 GCGGATGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 977
 QY 1741 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 DB 978 TGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1037
 QY 1801 GAGTGTGACAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 DB 1038 GAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1097
 QY 1861 GGGCATGCTGACAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 DB 1098 GGGCATGCTGACAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1157
 QY 1921 ATCAACCGGCTGTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 DB 1158 ATCAACCGGCTGTGTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1217
 QY 1981 CCCAGTGTGACATTTGGGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
 DB 1218 GCTGAGAGTATCTTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1277
 QY 2041 TGTAAACCATTTGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 DB 1278 TGTAAACCATTTGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1337
 QY 2101 CAACATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
 DB 1338 CAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1397
 QY 2161 GGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
 DB 1398 GGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1457

OY 2221 TACTGCACTGAGATGTCCTC 2242
 1 11111 111 111
 DB 1458 TTCTGCACACAGCCGTAGACCCC 1479

RESULT 5

AF444274 4290 bp mRNA linear ROD 06-DEC-2001
 DEFINITION Mus musculus Jedi protein mRNA, complete cds.
 AF444274
 VERSION AF444274.1 GI:17386052

SOURCE

Mus musculus.
 Mus musculus

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Mus.
 1 (bases 1 to 4290)
 Kivišov, A.V., Zinovjeva, M.V., Hendliks, J., Visser, J.W.M. and
 Belyavsky, A.V.

AUTHORS

TITLE

Jedi is a novel DSL and EGF-like repeat motif-containing protein
 expressed on non-differentiated hematopoietic cells

JOURNAL

2 (bases 1 to 4290)
 Kivišov, A.V., Zinovjeva, M.V., Visser, J.W.M. and Belyavsky, A.V.
 Submitted (07-NOV-2001) Stem Cell Biology, Lindsay F. Kimball
 Direct Submission
 Research Institute, New York Blood Center, 310 East 67 Street, New
 York, NY 10021, USA

JOURNAL

location/Qualifiers
 1..4290
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="E-7"
 /tissue_type="testis"
 138..3242
 /product="Jedi protein"
 /protein_id="AA138571.1"
 /db_xref="GI:17386053"
 /translation="MPLCPILLALGLRLTGLTNSNDPVNCFWESFTTTTRESHP
 FSLPAESCHRPEDHPCAPPVVRYRTYRVYKNDSPRIQCCGYTESKACVPL
 CAQCHVGRCAVAPNQCQAPRNGDCSECAPRMGPOCDKFCRGNNSGDKPKGA
 CTCPSSGLDPNCLQPCPAHYPAQCFDQCYGASCDPDGACFCFPGRAFSCKPA
 SOGTDGFCFPRTPYQMGVPGSGSCSPGPMGVISLPCPEGFHGPNTQECRC
 HNGGLDRTGQCHCAPYIGDRCQEECVGRGQCAATCCAPBARCFPANGACIC
 EHGFGDRTGCLCPGKXGLSCQECTCDPDRHSISCHMHGECSCQPMALHCHES
 CPODTHPGCCOHCICLHGCLGLADSLGRCAPIGTGPICANLCPDITGICSSKCS
 CENAIAGSPIDCTCKCEBGMKNSVPCPLGTWGNCAASQCAHADVCSPTQACT
 CTYPGHGACQLPCPGQFEGSCASVCDSDHSDGDPVHGRCQAGMGTGCHLPCP
 EGFWAGNSNTCTCKNGGTCSENGNCVCAFGFRPSCQRPFGYGRKVCQCKNN
 NMSCHPSDGTSCSLAGMTGPDSCBACPGHGLKSGOLCQCHGGTGCPDQSGICT
 PGMTPNCLGEGCPRMFGVNSOLCQCDGEMCHQOTGACVCPHSGADCMGSDS
 FTIMPTSPYTLGAVIGIATLVALIALFLFGYRQMKQKHEHLAVASTGRAL
 DGSQVMPDVPSPSYSHYSNRYHTLSQCSPPPNKVPVGSQLVSSQAPRPSRAH
 GRENHTYLPADMKHREPRERGASHLDNSYSCSYSHRNGPFGCHKGPISSEGLASV
 MSLSENVPATIRDLPSLPGERESGYEMKPPSVSPROSLHLRDRQROLOPORO
 SGTYPQSPFLSHNEESLSTPLPLPLPGYDSPKNSHIEPHYDLPVRRHPSPPSR
 RODR"

FEATURES

source

CDS

misc_feature
 402..524
 /note="Region: DSL domain"
 misc_feature
 528..2339
 /note="Region: contains 14 EGF-like repeat motifs"
 misc_feature
 2406..2465
 /note="Region: TM domain"
 misc_feature
 3087..3158
 /note="Region: PEST"
 BASE COUNT 816 a 1329 c 1156 g 988 t 1 others
 ORIGIN

Query Match 19.0%; Score 649.4; DB 10; Length 4290;
 Best Local Similarity 57.2%; Pred. No. 9.5e-189;
 Matches 1241; Conservative 0; Mismatches 916; Indels 12; Gaps .3;

OY 74 CTCTGAATCTTGAAAGACCTTAATGTGTGTAGACCACTGGGAAAGCTACTAGTGTGTC 133
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 190 CACTCAACTCCAAATGATGATCCCAATGTCTGTACTTTTGGGAAAGCTTCAACAGACACTA 249
 OY 134 AAGAGTCAATACCCACATCCCTTTGATCAAT---TTACTACAGAGCTGACATGATTC 190
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 250 AGGATGCCACCTCGCGCCCTTACAGCTGCTCCAGCTGAGATCTGCGCACAGGCCCTGG 309
 OY 191 TAACTGTTTAAATGACGCGGACAGAGACAGTACGTATGCGAGACCTTATGAGATGAGG 250
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 310 AGAGACCCCAACACTGTGTGCCAGCTTACGTTGTCTACGAGCTGTGACCTGAGGTGG 369
 OY 251 AGAAGACTATGTATATAGCGCAAGCTCAGTGTGCTGCTGATTTTATGAAAGCGGGAAA 310
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 370 TGAAGATGAGACTCCCGCCCAAGCTGAGTCTGTAGGGGTACTAGAGAGAGAGAGGG 429
 OY 311 TGTGTGCCCCCACTGTGTGTATTAATGTGTCCATGTGCTGTATTTGTCGAAACACT 370
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 430 CTTGTGTCACACTGTGTGCCAGAGAGTGTCTCATGTGTGCTGTGTGCTGCCAATCAGT 489
 OY 371 GTCAGTGTGAGCTGCTGGGAGGAGCAACACTGCTCAGTGCCTGGATGTGATCAGT 430
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 490 GCCAGTGTGACACAGGCTGGGAGGAGTGGCCAGCTGTCTCAGAGAGTGTGCCCGGAAAT 549
 OY 431 GGGGTCCCACTGACACAGCGGTGCCAGTGTGCAAAAATGGGCTGTGTGCAACCCATCA 490
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 550 GGGGACCCACAGTGTGACAAATTTGACACAGTGTGACCAACAAACAGTCTGTATGCCAAGA 609
 OY 491 CCGGGGCTTGCACACTGTGCTGCGGGGCTTCCGGGGCTGCGCTGAGAGACCCCTGTGAGC 550
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 610 GTGGGGGCTCTTTTGCCCTCTG6CTGAGCCCACTTACCTTACCTTGGCCCTG 669
 OY 551 AGGCGACATATGTGTAAAGACTGTATCAGAGATGCCAGTGTCCAGAAATGAGGACCTGCG 610
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 670 CCGGACCATATGTGTCTGCTGCGGAGTGTATTTGATTCAGTGTG---TATGGGACATCTGTG 726
 OY 611 ACCACGTCAAGGGGGAATGCGCTGCGCCACAGAGATACACGAGCCCTTGTGAGAGATC 670
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 727 ACCCCAGAGATGAGGCTCTTTCTGCGCTCAGAGAGAGAGACACAGCTGTAAATGTGC 786
 OY 671 TTTGTCTCTGTAAACATGATGTCAACAGTGTAGACAGAGATGCTTGTGAAAATGGAG 730
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 787 CCGTTCACAGGACACTGATGCTTTCTTGTGCCCCAGAACCTATCTTTGCAAAAATGGAG 846
 OY 731 GAGTGTATCATCACTGATCTGAGAAATGCTTTGCCCTTGTGCTGATGGGACAGTGT 790
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 847 GTTCTCTAGGGCTCCAAAGGCTCTGACAGTGTGCCACCGGCTGTGATGTGATTT 906
 OY 791 GTGTGAGGCTTCCCGGAGGCTGCTTTGGAAAGAACTGTTCCTCAAGATGCCAGTGC 850
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 907 GTTCCCTGCAATGCCAGAGGGGTTTCAATGAGACCAACTGTACTCGAATGTGCGTGC 966
 OY 851 AATATGAGGAGAGCTGTATGCTGCAACAGGCAATGTATGACAGTCCAGATACACAG 910
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 967 ACAAGGTGGCTCTGTATACAGTTTACTGCGGAGTGTCCACTGTGCTCTGCTATATG 1026
 OY 911 GGAAGCTGTCCAGAGATGATGCTCTGTGGAGACCTATGCGCTTCTGTGCTGAGACT 970
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 1027 GGGATCGGTGCCAAGAAAGAGTCCCGGTGGGCGCTTGTCAAGACTGTGTGAGACT 1086
 OY 971 GCCAGTGTCTCAAGGAGGAAAGTGTACACGTGAGCGGCCAATGCTGTGTGAAGCAG 1030
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 1087 GTGACTGTGCTCTGCGGCGCTTCTTCTTCTGTAATGGCCGCTGTGTGCGAAGCAGT 1146
 OY 1031 GCTTGTGCGGAGGCGCTGCAAGCAGCGCTGTGCTGAGGGGCTCTACGCAATCAAT 1090
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 1147 GCTTACAGGCGACGCTGTGACTGAGCGCTGTGCGGATGCGCTATGTGTGAGCT 1206
 OY 1091 GTGACAAAGGCTGTCTGCTGCACTTGTGAAACATCATATGATGTGACCCCATGTGTGGAG 1150
 1 11111 111 111 111 111 111 111 111 111 111 111 111 111 111
 DB 1207 GCCAGAGGCGCTGTGACCTGTGCGGACCAAGAAACACAGTCTACGTGCGACCCGATGACGGCG 1266

251 AGAAGATATATATAGCCGCAAGTCTCAGTGTCTCTGATTTTATGAAACGGGGAAA 310
 303 TGAAGAGGAGACTCCCGCCACGCTGCGATGCTAGGCGTTACTAGACAGAGAGGG 362
 311 TGTGTGCCCCACCTGTGTGATTAATGTGTGCTGCTGATGTATGCTCCAAACCT 370
 363 CCTGTGCTCCACTGTGTGCTGAGAGAGTGTCTCAATGCTGTGTGTGCTCCGATCAGT 422
 371 GTCAGTGTGAGCCTGTGCTGAGAGAGAGCAACTGTCTCAGTGTGCTGATGTATCAGT 430
 423 GCCAGTGTGACACAGGCTGAGGAGGCTGAGCAGTGTCTCAGAGTGTGCTCCGAGATGT 482
 431 GGGGTGCTCCACTGTGACCAAGCCGCTGAGTGTCAAAATGGGGCTGTGTGCAACCCATCA 490
 483 GGGGACACAGTGTGACAAATTTCTGCTGCTGAGCAACACAGTCTGTGTATCCCAAGA 542
 491 CCGGGCTTGTGCACTGTGTGCTGAGGCTTCCGGGGCTGCGCTGTGAGAGAGCGGTGAGC 550
 543 GTGGGGGCTGCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
 551 AGGGCAGCTATGTGTAACGAGTGTCTGATGAGATGCTGCTGCTGCTGCTGCTGCTGCTG 610
 603 CCGGCTCTATGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
 611 ACCAGTCAAGGGGGGAATGCGGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 670
 660 ACCCCAGAGATGAGCCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 719
 671 TTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
 720 CCGTGTCAAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 779
 731 GATGTGTCTACAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 790
 780 GTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 839
 791 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 850
 840 GTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 899
 851 ATATAGAGAGGAGTGTGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 910
 900 ACAAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 959
 911 GGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 970
 960 GGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1019
 971 GCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1030
 1020 GTGACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1079
 1031 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1090
 1080 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1139
 1091 GTATCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1150
 1140 GCCAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1199
 1151 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1210
 1200 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1259
 1211 TCTACGGGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1270
 1260 CGCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1319
 1271 TGACTGGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1330
 1320 ACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1379
 1331 CTCTGGAGACCTATGAGATAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390

Db 1380 CACCGGACACTATAGGATACACTGTCCTCCGCTGCTCTGTGAAATATGCCATGCT 1439
 1391 GCTCTCTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450
 1440 GCTCTCTGTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1499
 1451 TCAGATGTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1510
 1500 TTCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559
 1511 GGGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1570
 1560 ATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619
 1571 ATGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1630
 1620 ACTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1679
 1631 GCAGCCAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1690
 1680 GTGACCACTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739
 1691 CAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1750
 1740 TGGGCAAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1799
 1751 CCGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1810
 1800 CCGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1859
 1811 CAGCTTCCGAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1870
 1860 CAGGCTTCCGAGGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1919
 1871 GCAGCCAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1930
 1920 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1973
 1931 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1990
 1974 CCGTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2033
 1991 GATTGGGAGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2050
 2034 ACTGGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093
 2051 TTGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2110
 2094 AGGATGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2153
 2111 CACTGCTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2170
 2154 CACCAAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2213
 2171 GCAGCGCTTACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2227
 2214 GCCACCCACAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2270

RESULT 7
 AF440279 4482 bp mRNA linear ROD 20-NOV-2001
 LOCUS Mus musculus MEGF12 (Megf12) mRNA, complete cds.
 DEFINITION AF440279
 ACCESSION AF440279.1 GI:17017250
 VERSION
 KEYWORDS
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 4482)
 Ivanova, N.B. and Lemischka, I.R.

1511 GGGGACCTGCAACACCTTGAAGGAGCTGACAGTGTGACCTGTGATGGCGGGAGCA 1570
 1827 ACGAGTCTGACAGCCCAAACTGAGAGCTGTGACCTGTGACCTGTGATGGCGGGAGCA 1886
 1571 AATGCAATCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1630
 1887 ACTGCAAGTCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1946
 1631 GCAGCAGCAGATGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1690
 1947 GTGACCACTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2006
 1691 CAGGTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1750
 2007 TGGGCAACAGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2066
 1751 CCGTCTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1810
 2067 CCGTCTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2126
 1811 CAGGTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1870
 2127 CAGGTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2186
 1871 GCAGCAGCAGATGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1930
 2187 GTGCTGCA-----ATGCAAGTGTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2240
 1931 TGTGTACTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1990
 2241 CCGTCTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2300
 1991 GATTGGGCAAACTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2050
 2301 ACTGGGAGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2360
 2051 TTGACAGATCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2110
 2361 AGATGGAGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2420
 2111 CAGCTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2170
 2421 CAGCTGCTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2480
 2171 GCAGCAGCAGATGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2230
 2481 GCAGCAGCAGATGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 2540
 2231 AGAGATGCT 2239
 2541 TGGGAGCC 2549

RESULT 8
RATORD
LOCUS Ratord 660 bp mRNA linear ROD 09-AUG-1996
DEFINITION Ratord norvegicus (clone REM4) ORF mRNA, partial cds.
VERSION L41686.1 GI:780366
KEYWORDS monoclonal autoantibody;
SOURCE Ratord norvegicus (strain Holzman) cDNA to mRNA.
ORGANISM Ratord norvegicus
REFERENCE 1 (bases 1 to 660)
AUTHORS Asakura, K., Pogulis, R.J., Pease, L.R. and Rodriguez, M.
TITLE A monoclonal autoantibody which promotes central nervous system remyelination is highly polyreactive to multiple known and novel antigens
JOURNAL J. Neuroimmunol. 65 (1), 11-19 (1996)
MEDLINE 96235155

PUBMED 8642059
FEATURES
source location/Qualifiers
 1..660
 /organism="Rattus norvegicus"
 /strain="Holzman"
 /db_xref="taxon:10116"
 /clone="REM4"
 /issue="type="brain"
 /dev_stage="neonate"
 <1..>660
 <1..>660
 /note="ORF"
 /codon_start=1
 /protein_id="AAB05844.1"
 /db_xref="GI:780367"
 /translation="Ratord norvegicus (clone REM4) ORF mRNA, partial cds."

BASE COUNT 141 a 160 c 205 g 154 t
ORIGIN
 Query Match 16.0%; Score 548.2; DB 10; Length 660;
 Best Local Similarity 89.6%; Pred. No. 1.4e-157;
 Matches 589; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

598 GAGGACCTGCAACACCTTGAAGGAGCTGACAGTGTGACCTGTGATGGCGGGAGCA 657
 4 GAGGACCTGCAACACCTTGAAGGAGCTGACAGTGTGACCTGTGATGGCGGGAGCA 63
 658 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 717
 64 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 123
 718 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 777
 124 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 183
 778 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 837
 184 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 243
 838 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 897
 244 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 957
 898 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 957
 304 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 363
 958 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1017
 364 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 423
 1018 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1077
 424 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 483
 1078 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1137
 484 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 543
 1138 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1197
 544 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 603
 1198 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 1254
 604 TTCTGTGAGATCTTGTGCTGCTGCAAGATGCAAGTACGGGCTGAACTGTGCTGACCGTGGACT 660

RESULT 9

AX079681
 LOCUS AX079681 632 bp DNA Linear PAT 22-FEB-2001
 DEFINITION Sequence 425 from Patent WO0107611.
 ACCESSION AK079681
 VERSION AK079681.1 GI:13159250
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 632)
 AUTHORS Baker, K.P., Goddard, A. and Wood, W.I.
 TITLE Human polypeptides and methods for the use thereof
 JOURNAL Patent: WO 0107611-A 425 01-FEB-2001;
 Genentech, Inc. (US)
 FEATURES
 Location/Qualifiers
 source 1..632
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 BASE COUNT 145 a 159 c 172 g 154 t 2 others
 ORIGIN
 Query Match 15.6%; Score 534.8; DB 6; Length 632;
 Best Local Similarity 95.3%; Pred. No. 2e-153;
 Matches 593; Conservative 0; Mismatches 24; Indels 5; Gaps 4;

LOCUS AK074121 4470 bp mRNA Linear PRI 15-FEB-2002
 DEFINITION Homo sapiens mRNA for FLJ00193 protein.
 ACCESSION AK074121
 VERSION AK074121.1 GI:18676589
 KEYWORDS fls (full insert sequence).
 SOURCE Homo sapiens adult spleen cDNA to mRNA, clone:FLJ00193.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1
 AUTHORS Jikuya, H., Takano, J., Nomura, N., Kikuno, R., Nagase, T. and Ohara, O.
 TITLE The nucleotide sequence of a long cDNA clone isolated from human spleen
 JOURNAL Published Only in Database (2002)
 REFERENCE 2 (bases 1 to 4470)
 AUTHORS Jikuya, H., Takano, J., Nomura, N., Kikuno, R., Nagase, T. and Ohara, O.
 TITLE Direct Submission
 JOURNAL Submitted (21-JAN-2002) Takahiro Nagase, Kazusa DNA Research Institute, Department of Human Gene Research; 1533-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnaif@kazusa.or.jp, URL: http://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913, Fax:81-438-52-3914)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert and 5'- & 3'-end one pass sequencing; Research Association for Biotechnology; cDNA library construction and clone selection; Kazusa DNA Research Institute.
 FEATURES
 Location/Qualifiers
 source 1..4470
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="FLJ00193"
 /tissue_type="spleen"
 /dev_stage="adult"
 /note="vector: pBluescriptII SK plus"
 gene 1..4470
 /gene="FLJ00193"
 /transcript="1367..2842"
 /gene="FLJ00193"
 /note="Start codon is not identified."
 /codon_start=1
 /evidence="not_experimental"
 /product="FLJ00193 protein"
 /protein_id="BAB84947.1"
 /db_xref="GI:18676590"
 CDS
 /translation="PCSWTLVDFGFMGKARHLSCEBIMKVCNSNCTCANGSTCLP
 ENGNVCAPGFRGSPSCQSRGGRKRCVPCCKANHSFCHSGTCCLAGWGPDC
 SQQPPGHMGENCATOCQCHNGGTCHPDQSGICPLGWTGHHCEGCLGTFGANCQ
 POCSPGKCHPRTGACVCPHSGAPCRIGIOEPFTVPTPAVNLGAVIGIAYL
 GSLVALVALPIGYRHMOKGKEHHNLAVASSGRLDGSEYVMPVPYSYNSPSY
 HTLSQSPNPPRPKRVGRPLDASLQNPFRPGAGQHDHTLLPDMKIRRRPPRPGLD
 RGSRLDRSYSTYSNGPFRYKGLISEELGASVASLSENPATIRDLPSLPGP
 RESSYMEKGPSPSPRPQPPQDFMDQRRQPPQRRQSGTEQPSPLIHDSVGSOP
 PLPPLPGHDSKRNSHIRPHYDLPPVRHPPSPPLRRQDR"
 BASE COUNT 888 a 1372 c 1240 g 970 t
 ORIGIN
 Query Match 13.3%; Score 456.8; DB 9; Length 4470;
 Best Local Similarity 55.5%; Pred. No. 4.4e-129;
 Matches 1067; Conservative 0; Mismatches 767; Indels 90; Gaps 6;

541 CGCTGTAGCAGGCGACCTATGTAACGACTGTATCAGAGATGCCAGTCCAGATGGA 600
 181 CCCCTGACCCCTGCTACTATGAGCCCTGCTCCGACATGCTCCGCTCCAGTGC---ATGGG 237
 601 GCCACCTGCGACACGTCAGAGGGGAAATGCGCGTCCACAGAGATACACCGAGCCTTC 660
 238 GCACCCCTGCAATCCCGACACTGAGAGCTGCTTGTGCCCCGAGAGAGAACCTGGCCAGC 297
 661 TGTAGAGATCTTGTCTCTCTGTAACATGATGTCACAGTGTGAGCAGAGATGCTT-G 719
 298 TGTAGAGATCTTGTCTCTCTGTAACATGATGTCACAGTGTGAGCAGAGATGCTT-G 719
 720 TCAAAATGAGAGAGTGTGTATCATCAGTACATGAGAAATGCTCTTGGCTTGTGGTAT 779
 358 CCAAAATGAGAGAGTGTGTATCATCAGTACATGAGAAATGCTCTTGGCTTGTGGTAT 779
 780 GGGCAGCAGTGTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTT 839
 418 GGGCAGCAGTGTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTT 839
 840 ATGCCAGTGTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTTGTGAGCCTT 899
 478 ATGTGCTGCTGACACACGCGGCTCTGTGACACGATGAGGAGTGTGAGCCTTGTGAGCCTT 537
 900 AGGATACACAGGGGAAAC----- 916
 538 GGGTACACTGGGATGCTGAGTGGCTGGGCGGGGAGAGAGGAGAGAGAGAGAGAGAGAGAG 597
 917 -----GATCCAGAGATGATGCTGCTTGGGA 943
 598 GGGCGCGCGGACAGGCTACAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 657
 944 CCAATGCGCTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1003
 658 GCTTGTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 717
 1004 TGAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1063
 718 CCAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 777
 1064 GTCCTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1123
 778 GCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 837
 1124 CTATAGCTGTACCCCATCTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1183
 838 GCTTGTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 897
 1184 ACTGTATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1243
 898 ACTGTATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 957
 1244 GCCAAATGAGGCGAGCTGTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1303
 958 GCTTGTGGGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1017
 1304 AGGAAATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363
 1018 CCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1077
 1364 GCTGTGCT 1423
 1078 GCTGTGCT 1137
 1138 AAGGTGCT 1197
 1484 GTAATGCT 1543
 1198 GCAATGCT 1257
 1544 CGTGTGCT 1603

1258 CTTGCAACCCCTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1317
 1604 GCTGCAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1663
 1318 GAGAGGTTGCT 1377
 1664 GCTATGCT 1723
 1378 GAGCCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1436
 1724 GAGCCTGCT 1783
 1437 GCTATGCT 1496
 1784 CTATGCT 1843
 1497 CTGAGATGCT 1556
 1844 GCT 1903
 1557 GTCAGCT 1610
 1904 GCGGCGCT 1663
 1611 CC---TTCCTGCT 1667
 1964 TCTGCAATGAGTGTGCT 2023
 1668 ACTGCT 1727
 2024 GCAACCAACAGGAGCT 2083
 1728 GTCACCATGCT 1787
 2084 TTGGCAGTGTGCT 2143
 1788 CTGAGACACCTGCT 1847
 2144 CGTCAACCTGCT 2203
 1848 CATGCCAGTGTGCT 1907
 2204 CTGG 2207
 1908 CAGG 1911

RESULT 11
 AB011532
 LOCUS
 DEFINITION
 Rattus norvegicus mRNA for MEF6, complete cds.
 AB011532
 AB011532.1 GI:3449293
 VERSION
 MEF6
 SOURCE
 Rattus norvegicus (strain: Sprague-Dawley) adult brain cDNA to mRNA,
 clone 11b.pSPORT 1 clone: RG1805.
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE
 1 (sites)
 Nakayama, M., Nakajima, D., Nagase, T., Nomura, N., Seki, N. and
 Ohara, O.
 Identification of high-molecular-weight proteins with multiple
 EGF-like motifs by motif-trap screening
 Genomics 51 (1), 27-34 (1998)
 JOURNAL
 MEDLINE
 9836089
 2 (bases 1 to 5523)
 REFERENCE
 Nakayama, M., Nakajima, D. and Ohara, O.
 Direct Submission
 Submitted (26-FEB-1998) Manabu Nakayama, Kazusa DNA Research
 Institute, Laboratory of DNA technology; 1532-3, Iana, Kisarazu,

Chiba 292-0812, Japan (E-mail: mmanabu@kazusa.or.jp,
Tel:+81-438-52-3915, Fax:+81-438-52-3914)

FEATURES

SOURCE

Location/Qualifiers
1..5523

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="RG1805"
/tissue_type="brain"
/clone_1lb="psport 1"
/dev_stage="adult"

gene

1..5523

CDS

/gene="MEGF6"
235..4959
/gene="MEGF6"
/codon_start=1
/product="MEGF6"
/protein_id="BAA32462.1"
/translation="MPVRAEAAARVVALALLLPAMPASPPILPPIOPMPHYC
ABOKLTIHSHPOQAFRIYPIVPRIRGCAQQAQACIGQERRTYMYNROYTAEAR
TVRRCCPRHSQKROQEGCLSDYDECASANGCEBPCNTVGGFTYRCPPG10LQDGK
TCQDYDECAHNGCQHRVNTPGSYLCECKGFRLHTDRTCLAISSCTLGNGCQH
OCVOLTVQHRCQRPYQLOEDGRVRRSPCAENGSCMHICQELRG1AHCCHPG
YOLADRTKEDVDECA1G1AOCAGHCNTOGSEFCVCHAGYELGADROCIRIMEI
VNSCAGNGSGSHGSHSTGPTGLCTCPRGYELDEDKTICIDIDDCANSPPCCQACANT
PGYECSCFAGRLNTDGGCGEDVDECAHNGCEBPCNTVGGFTYRCPPG10LQDGK
RRCSTLSESYVDIDGRIPVPRPLPHIVLHDELRLTODDYGAEEBAABAE1REHT
LTERKVLDSHFGHDCSLTDCDRNGGTCFPOQDDCDPEEGTGLIC1CETCPD1FGK
NCSPTCQNGGTCDPLVACRCPVPGAHCEDECPKGFGRKHKRKHCHCNRGRH
RLYGALCDPLGLYGRFCHLACPPNAFGSCSEDCLEQSHTESCPNKDSCSKAGFQ
GERCOAECSGFFGPGCHRRCTCGPGVACDPVSGCRTOCPGYGEGCGGSCPPGTE
GVNSGSCSYGAPCHRYTGECLCPRGTEGEGADCEBGRMKGLCOPIELPACSEHAS
CNPERGTCLCPGYRSCODTCSNRTGTGTQITCAANDGHCPTTGRSCAGAWT
GLSCORACDSEHMGPDCTHPNCNSAGHNCDSVSLACLCEAGTEBPRCEQSCROGTG
PSEQRCKRCHGAADHVSAGACTCPAGNRGSCFEHACPAFGFGLDCDSACNSAAPC
DAVSESCIPAGRMGPRCAQSCPLTFLGNSQ1CTCFNGASCDSDVTGSC1CAPMG
PTLOACPGPLGYKNGHSC1CRNGRCDPLIGQCTCEPMTGLACENDECLGHAAQ
COLNCSLHGGLCDRLTGHCLCPAGMTDKOOSVSTFGVHCEHACRGAACH
VTGACFCPPMRGPHCEBACPRMGEMCAQRCLPTNASHCHVHYEBCRPFGTGLS
CBQACQPGTFCDBLDCQPEWTACDPASGVCTCAAGYHGTGCLQCRPSRYTPGC
EHIC1KNGTCDPATYACCPAGFLGADCSLACPOGRFPGCAHVCACROGACDPV
SGAC1SPKGTGVRCEHGCPODRFEGCELCACRNGGLCHATNNGSCPLGMMPHC
BHACPAHYGAC1LEFCRNNNGSC1EPTTGAC1LGPYGGAC1CHESPGFPCCOR
VCECOGAPDPVSGOCLCPAGFHGCRFEGKCEGSESPDGLQCCNCHTGVPCD1SG
LC1CPBRTGAAC1DCCRFRGPECALRCDGGGADCDP1SGQCHVDSTMGPTCRE
VPTQ1SSRPAPQHPSSRAMH"

polyA_site

5523

BASE COUNT 960 a 1603 c 1797 g 1163 t
ORIGIN
/gene="MEGF6"
/note="31 a nucleotides"
/note="11 a nucleotides"

Query Match 11.7%; Score 400.8; DB 10; Length 5523;
Best Local Similarity 51.3%; Pred. No. 1e-111;
Matches 1092; Conservative 0; Mismatches 1017; Indels 21; Gaps 6;

QY 372 TCAGTGTGAGCTGCTGGGGAGGAGACCAACTGCTCCAGTGCCTGGGATGATGATACG 431
DB 2394 TCAGTGTGAGCTGCTGGGGAGGAGACCAACTGCTCCAGTGCCTGGGATGATGATACG 2453
QY 432 GGGTCCCACTGACCAAGCCGGTGCAGTCAAAAATGGGGCTCTGTGCAACCCCATCAG 491
DB 2454 TGGTGTGAAGTGTCTGAGTTCCTGTTCTCTGC---GTTGAGAGCCCTTGTACCGAGTCA 2510
QY 492 CGGGGCTTGCACCTGTGCTGCGGGGCTTCGCGGGCTGGCGCTGGAGAGACCGCTGTGAGA 551
DB 2511 AGGGGAGTGTCTGTGCGCCACAGAGAACTGGGGAGAGTGGAGTGGAGTGTGCCA 2570
QY 552 GGGGACCTATGTAAGCATGTCTCAGAGATGCC---AGTGCAGAAATGAGACCCAGT 608
DB 2571 GGGGCTTGGGGCTGCTGCTCAAGATCTGCCCGCATGTGAGACAGGAGCCAGCTG 2630

QY 609 CGACACAGTCAAGCGGGGAATGCCGCTGCCACACAGGATACACCGAGCCTCTGTGAGA 668
DB 2631 CAACCTTGAGACTGGAACTGCTGTGTCTCCCTGGCTTTGTTGTAAGCCGCTGCCAGGA 2690
QY 669 TCTTGTCTCTCTGTGTAACATGCTGCACAGTGTGAGAGAGATGCTGTCAAAATG 728
DB 2691 CACTTCTCAGCAGGCTGTGAGGAGACTGGCTCCAGATAAGGTGTGCTTGTGCCAAGA 2750
QY 729 AGGAGTGTATCATCAGCTACGTGAGAAATGCTTTCCTTCTGCTGTGATGGGACAGT 788
DB 2751 TGGGCAATGTGACCAACACCGAGAGCTGTAGCTGTGCCCCGGGTGAGCTGCGAC 2810
QY 789 GTGTGTGACCTTGTGCCCGGAGGCTCTTGTGAAAGAACTGTTCCAAGAAATGCCAGT 848
DB 2811 CTGCCAGAGAGCTGTGACACATGTGACACTGGGGGCCCGACATGATCACCTTGCACAT 2870
QY 849 CCATATATG---AGGAGCTGTGATGCTGCCACAGGCCAATGATCATTTGACATCAGATA 905
DB 2871 CAGTGGGGCCATGGGAATTTGTAGCCCGTCAGTGGCTCTGCTGTGAGGCGCGCTA 2930
QY 906 CACAGGGGAAGGTGCGCAGAGATGATGTCTGTGGAGCTATGGCGTCTCTGTGCTGA 965
DB 2931 TGAAGGTCCACGCTGTGACACTGTGTGCGCAGGGAATCTATGAGGCCACAGCTGAGA 2990
QY 966 GACCTCCAGTGTGTCAACGAGGAGGAAGTTCACACGTAGCGGGCGCATGCTCTGTGA 1025
DB 2991 GAAGTCAATGTGACATGAGGAGCTGCTGTGACATGTTCAGTGGGGCTGTATACGTGCC 3050
QY 1026 AGCAGCTTGTCTGGGAGCGCTGCGAGACGCGCTGTGTCTGTGAGGGGCTTACGGCAT 1085
DB 3051 ACCCGGCTGAGGAGGAAGCTTGTGTGAACATGCC---TGTCCGTGCTGCTTCTTGGT 3107
QY 1086 CAAATGTCAAAAGGTGTCCCTGCCACTTGGAAAACACTCAATAGCTGTACCCCATGTC 1145
DB 3108 AGACTGTGACA-----GTGCTGCAACTGTCTGTGCGGGGCACTGTGATGTCTGCAC 3161
QY 1146 TGGAGAGTGTGCTGTGCAACCGGGCTGTGACAGACTCTGTGAATGAAACATGTTCTCC 1205
DB 3162 CGGCTCATGATGTGCGCAGCTGTGCGGCTGTGCGGCCACAGCTGTGCCAAGTTGCCAC 3221
QY 1206 TGGATTTACGGGGAAGCTTGTGCCAGAGATCTGACACTGCCAAATGGGGAGACTGTGA 1265
DB 3222 CTTACACTTGTGGGCTTAATGTGACAGCATGTTGTACTGTTCATAGGGGCTCTGTGA 3281
QY 1266 CAGTGTGACTGGAATGTGACACTGTGTGCCAGATTCAAAGAAATGATGTCTTACCCC 1325
DB 3282 CTCTGTCAACAGGCAATGTGACACTGTGTGCCCTGTGATGGAGCCACACTGCTGTGAG 3341
QY 1326 ATGCCCTGTGGGAACCTATGGATTAAGTGTCTGTGCTGTGGCTGTAAATATGATGC 1385
DB 3342 CTGCTCTGTGGCTGTATGTGCAAGAACTGTGACATTTCTGTCTTGTGCCGCAAGGAG 3401
QY 1386 AGTGTCTGTCTGTGAGAGGCTTGTGTACTGTGCAAGGAGAGCTGTGAGAGGAGTGTG 1445
DB 3402 GAGATGTGACCCATGTTGTGGGCGACAGTGCACAGTGCACAGAGGTTGAGCCGCTTGG 3461
QY 1446 CTTCATCAATGTGCCAGTGTGACATGTGGGCTTGTGCTGTAACTTAATGTCAGTGCCT 1505
DB 3462 TGAGAAAGAGTCTCTTGTGACACTATGTACAGCGGTTGTGAGCTCAACTGAGTTGCT 3521
QY 1506 CAAGGGGAGGAGCTGTGCAACACCTGTGAGCGGAGACTCAGCTGTGACACTGTGATGG 1565
DB 3522 TCACGGGGGCAATGTGACCGGCTTACCGGCCACTGTCTGTGCCAGCTGTGAGTGTG 3581
QY 1566 GGAAGAAATGCAATCTTCTGTGCCAGATGTGACAGTACGAGGCTGTGACTGTGAGCGCT 1625
DB 3582 GGACAAAGTGTACAGAGCTCTGTGCTGACAGTGTGAGGCTTGTGAGGAGTGTGAG 3641
QY 1626 CCAGTGTGACCCAGCAGATGTGCTGCACCTTACACAGGCGCAATGTGCGCTGTGCCGG 1685
DB 3642 TGTCTGTCCGAAAGGTGCTCTGTGACCATGTGACAGGAGGCTGTGCTGTCTGTGAG 3701
QY 1686 ATGTGTAGTGTTCACACTGTGACAGCTGTGTGTGAGGAGACCTGTGGGCGCCCAACTGCT 1745

OY	549	GCAGGCAACCAATGATGTAACAGACTGTATCATGAGATATCCAGATGCGAAGAAATGAGGCCACTG	608
Db	1685	CAAGGGCTACTACTATGGCAAGCACTGTGCAAGAAATCAACTGTGCCAACCGGGGGCCGGTG	1744
OY	609	CGACCAACGTCACGGGGGAATGACCGCTGTGCCACCAAGGATATCACCGGAGCCCTTGTGTGAGGA	668
Db	1745	CCACCGGCTCTACGGGGGCTGTGCTCTTGTGCAACCCAGGGGCTTACGGCCGCTTTGTCCACCT	1804
OY	659	TCTTTGTCCTCTGTGTAACATGATGATCCACAGTGTGAGCAGAGATGCCCTGT-----CA	722
Db	1805	CACCTGCCCGCCGCGGGCCCTTTGTGGGCCGGGGCTGTGAGGAGAGGTGCAGATGTGTGCACGC	1864
OY	723	AAATGAGAGAGTGTGTATATACGCTACTGGAGATCTCTTGCCCTTGTGCTGCATGGG	782
Db	1865	CCACACGCACTCTGTGTACAAAGAGGATGTGCACGCTCTCTCAAGAGGCTGTCCGGG	1924
OY	783	CACAGTGTGTGTGAGGCTTGCCCGGAGGTGCTTTGGAAAGAACTGTATCCAGAGATG	842
Db	1925	CGACGCTGTGTACGCAAGATGTGAGCTGGGCTTACTTGTGAGCCGGGGGTGTGGCAGGACATG	1988
OY	843	CCAATGCCATTAATGAGAGGAGCGTGTATGTCTGTCCACAGGCCAATGT-----CATTTGAG	896
Db	1985	CACCTGCCCATGTGGGCGCTGTGGCTGTGTACTGCTCGTAGCGGCGAGTGTGGGAACGGGTGTCC	2044
OY	897	TCCAGGATACACAGGGGAACGGTGTCCAGATGATGTCTCTTTGGGACCTATGCGCTTCT	956
Db	2045	TGCTGGGCTTCCAGGGAAGACACTGTGGCCAAAGATCCCGGTGTGGGACTTTTGGCGTGA	2104
OY	957	CTGTGCTGAGACCTTGCCAGTGTGTCAACGAGAGGAAAGTATTACACGTGAGCGGGCGATG	1016
Db	2105	TGTGCTCGAGCTCTGTCTCTGTGTG-----GGGGGGCCCCCTCCACCGGGGTGTACGGGGCAGTG	2161
OY	1017	CCCTGTGGAAGCAGGCTTTTGTCTGTGGCGAGCGCTGTGAGACGCGCTGTGTCTGAGAGGGCT	1076
Db	2162	CCGGGTGTCAACCGGGGAGGACGTGGGGGAACGTGTGAGGCA---GATTGTCCGGAAGGGCGG	2218
OY	1077	CTACGGCATCAATATGTACAAACGGTGTCCCTGTGCCACTTGGAAAACCTCATAGCTGTCA	1138
Db	2219	CTGGGGGCTGTGGGCTGCCAGAGATGTGTGCC---AGCATCCACAGACGCTGCCCGCTGGGA	2275
OY	1137	CCCCATGTCGTGAGAGTGTGTCCGTCAAGCGCGGCTGTGTGAGACATCTATGTAAGAAC	1196
Db	2276	CCCTGTAGACCGGAGCCTGTGCTGTGCCCTTCCCTGTGGCTTGTGTGGGAGCGGTGTCCAGAGCT	2335
OY	1197	ATGTTCCTCTGTGATTCTTACCGGGGAAGCTTGTGCCAGCATGTCTACAGCTGTCCAAATATGGGAGC	1256
Db	2336	GTCGCCACGACGGGTGTGTATGTGTCCAGGTGTCCAGAACAAAGTGTCTGTGTGCCAATGATGTG	2395
OY	1257	AGACTGTACAGTGTGACTGTGAAAATGTACACTGTGTGCCCGCAGGATTTCAAAGAAATTGACTG	1318
Db	2396	GCACTGTGCCACCCACCGCACCCGACACTGTGACGCTGTGCCCCGGGGTGTGAACCGGCTTTAGCTG	2455
OY	1317	CTTACCCCATGCGCCCTGTGGGAACGTAATGGAATTAACATGTCTCTCGGTGTGGCTGTAA	1376
Db	2456	CCAGAGAGCCTGTGTATCTGTGGCACTGTGGGAGCCTGTGACTACAGCACCCCTGTCAACTGTGAG	2515
OY	1377	AAATG---ATGCAAGTGTGTCTCTCTGTGTGAGCGGGTGTGTACTTGTCAAGGACGCTGTGCA	1433
Db	2516	CGCTGTGGCAAGGAGGTGTATGTGCATACAGCGGCTGTGTCTGTGTGAGAGCTGTGCTTACTGT	2575
OY	1434	CGGGGTGAGCTGCTCATCATGATGTGTCCACAGTGGCACATGGGGCTTTGTGCTTAACTTAAC	1493
Db	2576	GGGGCCCCGGGTGTGAGCAGCAGTGTGTCCCAAGGGCCACTTTTGGGCCGGGTGTGAGCAGCT	2635
OY	1494	ATGCGCATGTGCTCAACGGGGGAGCCTGTGCAACACCTGTGAGACGGGACCTGTGACGTGTGTACAC	1553
Db	2636	GTCGCACGTGTACGATGGAACACACCTGTGTGACACGTCACACGGGGGCTGTGACCTGTGCCGGC	2695
OY	1554	TGATGTGCGGGGGAAGAAATGTCCATCTTCCCTCCAGATGTGCACAGTATGAGGCTCAACTG	1613
Db	2696	CGGCTGTGAGGGGCACTTTTGTGCGAGCATGTCTGTGCCGGCGGCTCTTTTGGATTGTGACTG	2755
OY	1614	TGCTGAGCGCTGTGCATGTACAGCCACGACAGATGGCTGTCCACCTTACACAGGGGCAATTGGCG	1673

Dd	2756	TTCGAGTCCCTGCACAACTGCACCGCCGGAGCTGCCGTGTATGCGCGTAATGGCTCCTGCT	2815
Oy	1674	CTGCCTCCCGGGATGTGATGAGTGTTCACTGATGACAGCGTGTGTGCTGAAGGACGCTGGGG	1733
Dd	2816	CTGCCCCGCTGGCCCGCGGGGCCCCCCTGCTGCCAAGAATGCCCTCCGCCGGAGCTGAG	2875
Oy	1734	CCCCACTGCTCCCTGCTGCTACTGTATAAAAATGGGCTTCATGCTCCCTCATGATGATGG	1793
Dd	2876	AGCTTGCTGCGCGGCACACGCGCGGTGGCTCAACCGGGGGCCGTGTGACCGGCACACGGG	2935
Oy	1794	CATCTGCGAGTGTGACCAGGCTTCCGAGGCAACCACTGTACAGAGATCTGTCCCTGG	1853
Dd	2936	CCGCTGCTCTGCGCCAGCGGGCTGGACTGGGGACAAGTGTACAGACCCCTCTGCGGGG	2995
Oy	1854	TTTTATATGGGATGCTGCTCAGACACACATGCGCACAGTGCCTTCACAGACGGCGGCTG	1913
Dd	2996	CTGGTTTGGAGAGGCGCTTGCCCAAGCGCTG-----AGTTCGCCGCTGCCGCTGCTG	3049
Oy	1914	CCACCACATCAACGGGCTGTGTGACTGCTGGCTGGCTTCACAGGCGCCCTCTGCATGA	1973
Dd	3050	CCACCAACGTCACTGGGGCGCTGGCGGCTGCCCCCCCTGGCTTCACTGCTCGCGGCTGGAGCA	3109
Oy	1974	AGTGTGTCCCAATGGCAATTGGGAAAACTGTGTACAGAAATTTTACTGTACCAACAA	2033
Dd	3110	GGGAGTGTCCGCGCGGGCGGTATGGCCAGGCTGTATACAGCTGTGTGGTGTCTTCACGG	3169
Oy	2034	CGGAACTGTAAACCCCATTTGACAGATGTTGTACAGTGTATACCCCGGTGTGATTTGGCAGTA	2093
Dd	3170	GGGCTCTGTGTATGTCGGCCAGGGGGCCCTGCCGCTGCCCACTGGGTTCTTCGGAGCGGA	3229
Oy	2094	CTGCTCTCAACCATGTCCACTGCGCCACTGGGGCCCAAATGTCAATCAACAGGTGCMACTG	2153
Dd	3230	CTGCACAACCTCAACCTGTCCGCAAGGGCGCGTGGCGCCCAATGTCAACCAAGTGTGGGTG	3289
Oy	2154	CCATATATGAGAGCTTCTGTGACGCGCTACGATGTGGGAAATGTAAATGCACTCCTGTGGTGAN	2213
Dd	3290	TGGGCAAGGGGGGGGCTGTGACCCCTGTGACCGGCAACCTGCTGTGCCCCCGGGGAGAGC	3349
Oy	2214	AGGGCTTACGTGCACTCAGAGATGTGCTTAGGGTTTATGGAAMAAAGATTTGTGCACTGAT	2273
Dd	3350	CGGCGTCCGCTGTGTAGCGACAGGCTGCCCCAGAACCGGTTTGGCTGTGAGGCTGCGAGACAC	3409
Oy	2274	ATGCCAATGTCAAAAAGGAGCTGACTGTGCGACACATTTCTGGGCGAGTACTTGTGCGCAC	2333
Dd	3410	CTGCTCTCGAGAAATATGGGGGCTGTGCGACGCGCAAGACGCACTCTCTCTGTGGGCT	3469
Oy	2334	TGGATTCATGGGACGGCACTGTGACACAGTAGTCCCTTCAAGAACATATGTGCTATGCTG	2393
Dd	3470	GGGCTGTGAGCGGGGCGGCACTGTGACACTGGCTGTGCCCTTGGCGCTACGGAGCGGCTG	3529
Oy	2394	TGCGCAGATATGTGATTTCTMTGGAACAACATCCAGCCGCGACACACATCACTGGGAGCTGTA	2453
Dd	3530	CCATTTGAGATGTCTCTCCACACACACAGCAGCTGTAGACCTGTCCACGGGCACTTGCGG	3589
Oy	2454	CTGCAGCCCCGAGTGAAGGAGCGAGATTTGATCA	2489
Dd	3590	CTGGCGCCCGGCTTCTATGGCAGGCGCTGGGAGGA	3625
<hr/>			
RESULT 13	AC026800/c	146206 bp DNA linear PRI 08-JUN-2001	
LOCUS	AC026800	Homo sapiens chromosome 5 clone CTD-2377N4, complete sequence.	
DEFINITION	AC026800		
VERSION	AC026800.4	GI:14329105	
KEYWORDS	HTG.		
SOURCE	Homo sapiens.		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 146206)		
	DOE Joint Genome Institute and Stanford Human Genome Center.		

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (28-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 146206)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (08-JUN-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jun 8, 2001 this sequence version replaced gi:13470154.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
WT-4499 G02896
SHGC-84555 G51780
SHGC-101777 G56263.
FEATURES
source 1.146206 Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2377N4"
BASE COUNT 46774 a 28468 c 27627 g 43337 t
ORIGIN
Query Match 7.3%; Score 250.8; DB 9; Length 146206;
Best Local Similarity 92.3%; Pred. No. 6.3e-65;
Matches 264; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 412 GCGTGGAGGATGATGACGAGGGGCTCCCACTGACACGCGGTCGACAGCAAAATGGG 471
DB 96376 GCGTGGAGGATGATGACGAGGGGCTCCCACTGACACGCGGTCGACAGCAAAATGGG 96317
QY 472 GCTCTGTGCAACCCATCACCAGGGGCTTCCACTGTCTGCGGGCTTCCGGGGCTGGCC 531
DB 96316 GCTCTGTGCAACCCATCACCAGGGGCTTCCACTGTCTGCGGGCTTCCGGGGCTGGCC 96257
QY 532 TCGGAGAGCCGCTGTGAGCAGGGGACCTATGTTACGACTGTCTATCAGAGATGCCAGTGC 591
DB 96256 TCGGAGAGCCGCTGTGAGCAGGGGACCTATGTTACGACTGTCTATCAGAGATGCCAGTGC 96197
QY 592 CAGATGAGAGCCATCGACGACGACGAGGGGAGATGCCCTGCCACACAGATACACC 651
DB 96196 CAGATGAGAGCCATCGACGACGACGAGGGGAGATGCCCTGCCACACAGATACACC 96137
QY 652 GGAGCCTTCTGTGAGATCTTGTCTCTCTGTGTAACATGATGCAC 697
DB 96136 GGAGCCTTCTGTGAGATCTTGTCTCTCTGTGTAACATGATGCAC 96091
RESULT 14
AC008566 152765 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTC-540P9, WORKING DRAFT SEQUENCE,
36 unordered pieces.
AC008566
AC008566.4 GI:13699384
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS 1 (bases 1 to 152765)
TITLE DOE Joint Genome Institute.
JOURNAL Sequencing of Human Chromosome 5
REFERENCE 2 (bases 1 to 152765)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7708954.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 392909, H410
Center clone name: CIT-HSPC_540P9

Summary Statistics
Consensus quality: 121755 bases at least Q40.
Consensus quality: 133065 bases at least Q30
Consensus quality: 136597 bases at least Q20
Estimated insert size: 120000; pulse field gel estimation
Estimated insert size: 149265; sum-of-coverage estimation
Quality coverage: 6.44 in Q20 bases; pulse field gel estimation
Quality coverage: 5.18 in Q20 bases; sum-of-coverage estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1072: contig of 1072 bp in length
* 1073 1172: gap of unknown length
* 1173 2239: contig of 1067 bp in length
* 2239 2340: gap of unknown length
* 2340 3463: contig of 1124 bp in length
* 3463 3564: gap of unknown length
* 3564 4791: contig of 1228 bp in length
* 4791 4891: gap of unknown length
* 4891 5928: contig of 1037 bp in length
* 5928 6029: gap of unknown length
* 6029 7174: contig of 1146 bp in length
* 7174 7275: gap of unknown length
* 7275 8490: contig of 1216 bp in length
* 8490 8590: gap of unknown length
* 8590 9611: contig of 1020 bp in length
* 9611 9710: gap of unknown length
* 9710 10819: contig of 1109 bp in length
* 10819 10820: gap of unknown length
* 10820 10920: contig of 1164 bp in length
* 10920 12083: contig of 1124 bp in length
* 12083 12183: gap of unknown length
* 12183 13307: contig of 1124 bp in length
* 13307 13407: gap of unknown length
* 13407 14635: contig of 1228 bp in length
* 14635 14735: gap of unknown length
* 14735 14736: gap of unknown length
* 14736 15839: contig of 1104 bp in length
* 15839 15939: gap of unknown length
* 15939 16982: contig of 1043 bp in length
* 16982 17082: gap of unknown length
* 17082 18125: contig of 1043 bp in length
* 18125 18225: gap of unknown length
* 18225 18226: gap of unknown length
* 18226 19372: contig of 1147 bp in length
* 19372 19373: gap of unknown length
* 19373 19473: contig of 1007 bp in length
* 19473 20479: gap of unknown length
* 20479 20580: contig of 1007 bp in length
* 20580 21587: gap of unknown length
* 21587 21686: gap of unknown length
* 21686 23352: contig of 1666 bp in length

```

* 23353 23452: gap of unknown length
* 23453 24736: contig of 1284 bp in length
* 24737 24836: gap of unknown length
* 24837 26369: contig of 1333 bp in length
* 26370 26469: gap of unknown length
* 26470 27511: contig of 1042 bp in length
* 27512 27611: gap of unknown length
* 27612 29306: contig of 1695 bp in length
* 29307 29406: gap of unknown length
* 29407 31230: contig of 1824 bp in length
* 31231 31330: gap of unknown length
* 31331 33290: contig of 1960 bp in length
* 33291 34562: contig of 1172 bp in length
* 34563 34662: gap of unknown length
* 34663 36347: contig of 1685 bp in length
* 36348 36447: gap of unknown length
* 36448 37601: contig of 1153 bp in length
* 37601 39099: contig of 1399 bp in length
* 39100 39199: gap of unknown length
* 39200 41630: contig of 2431 bp in length
* 41631 41730: gap of unknown length
* 41731 47766: contig of 6036 bp in length
* 47767 47866: gap of unknown length
* 47867 57350: contig of 9484 bp in length
* 57351 57450: gap of unknown length
* 57451 73202: contig of 15752 bp in length
* 73203 73302: gap of unknown length
* 73303 83756: contig of 10454 bp in length
* 83757 83856: gap of unknown length
* 83857 114166: contig of 30310 bp in length
* 114167 114266: gap of unknown length
* 114267 152765: contig of 38499 bp in length.
Location/Qualifiers
source
1..152765
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-540P9"
/clone_lib="Caltech human BMC library C"
BASE COUNT 43611 a 29766 c 30113 g 45764 t 3511 others
ORIGIN

```

```

Query Match 7.3%: Score 250.8; DB 2: Length 152765;
Best Local Similarity 92.3%: Pred. No. 6.3e-65;
Matches 264; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

```

DB 412 GCCTGGAGTGTGATCACTGGGGTCCCACTGCACCGCGGTGCCAGTGCATAAATGGG 471
GCCTGGAGTGTGATCACTGGGGTCCCACTGCACCGCGGTGCCAGTGCATAAATGGG 111093
DB 472 GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 531
GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
DB 111094 GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
DB 532 TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 591
TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
DB 111154 TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
DB 592 CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 651
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
DB 111214 CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
DB 652 GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
DB 111274 GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697

```

```

RESULT 15
AC010415/c AC010415 175144 bp DNA linear HTG 18-JUL-2000
LOCUS AC010415 175144 bp DNA linear HTG 18-JUL-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-217111, WORKING DRAFT SEQUENCE,

```

```

6 ordered pieces.
AC010415.5 GI:9256209
HTG: HTGS_PHASE2; HTGS_DRAFT.
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 175144)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 175144)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710704.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 693406
Center clone name: CTD-H1_217111
-----
Summary Statistics
Consensus quality: 161499 bases at least Q40
Consensus quality: 171400 bases at least Q30
Consensus quality: 172941 bases at least Q20
Estimated insert size: 173000; pulse field gel estimation
Estimated insert size: 174944; sum-of-contigs estimation
Quality coverage: 5.76 in Q20 bases; pulse field gel estimation
Quality coverage: 5.69 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a "working draft" sequence. It currently
* consists of 6 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1
76295 76394: contig of 76294 bp in length
76295 76395: gap of unknown length
76395 95691: contig of 19297 bp in length
95691 95791: gap of unknown length
95791 95797: contig of 3006 bp in length
95797 98798: gap of unknown length
98798 98898: gap of unknown length
154073 154173: gap of 55176 bp in length
154173 154174: gap of unknown length
154174 168054: contig of 13881 bp in length
168054 168155: gap of unknown length
168155 175144: contig of 6990 bp in length.
Location/Qualifiers
source
1..175144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-217111"
/clone_lib="Caltech human BMC library D"
BASE COUNT 55980 a 34135 c 32899 g 51583 t 547 others
ORIGIN

```

```

Query Match 7.3%: Score 250.8; DB 2: Length 175144;
Best Local Similarity 92.3%: Pred. No. 6.5e-65;
Matches 264; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

```

```

DB 412 GCCTGGAGTGTGATCACTGGGGTCCCACTGCACCGCGGTGCCAGTGCATAAATGGG 471
GCCTGGAGTGTGATCACTGGGGTCCCACTGCACCGCGGTGCCAGTGCATAAATGGG 111093
DB 472 GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 531
GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
DB 111094 GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
GCTCTGTGCAACCCATCAACCGGGGCTTGGCCACTGTGCTGGGGGCTTCCGGGGCTGGCG 111153
DB 532 TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 591
TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
DB 111154 TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
TGGGAGACCGCTGTGAGAGGGGACCTAATGTAACGACTGTATAGAGATGCCAGTGC 111213
DB 592 CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 651
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
DB 111214 CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
CAGAAATGAGACCACTGCACCACTGCACGGGGGATGGCGGTGCCACAGAGATACACC 111273
DB 652 GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
DB 111274 GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697
GAGAGCTTCTGTGAGAGATCTTTTCTCCTCGTGAACAATGCTGCAC 697

```

```

OY 472 GCTCTGTGCAACCCCATCAGCGGGGCTTGCCACTGTGCTGGGGCTTCGGGGCTGCGGC 531
    |||||||
Db 94590 GCTCTGTGCAACCCCATCAGCGGGGCTTGCCACTGTGCTGGGGCTTCGGGGCTGCGGC 94531
    |||||||
OY 532 TGGGAGGACCGCTGTGAGCAGGAGCCTATGTAAAGACTGTCAATCAGAGATGCCAGTGC 591
    |||||||
Db 94530 TGGGAGGACCGCTGTGAGCAGGAGCCTATGTAAAGACTGTCAATCAGAGATGCCAGTGC 94471
    |||||||
OY 592 CAGAAATGAGGACCACTGCGACACGTCACGGGGGAATGCGCTGCCACACAGATACACC 651
    |||||||
Db 94470 CAGAAATGAGGACCACTGCGACACGTCACGGGGGAATGCGCTGCCACACAGATACACC 94411
    |||||||
OY 652 GGAGCCTTCTGTGAGGATCTTTGTCCTCCTGTAAACATGGTCCAC 697
    |||||||
Db 94410 GGAGCCTTGTAAATCATGCTGCGCCAGCAGCAGAGAGCCAC 94365
    |||||||

```

Search completed: May 9, 2003, 10:55:18
 Job time : 10719 secs

GenoCore version 5.1.4.D5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 04:47:55 ; Search time 4559 Seconds

(without alignments)
12159.937 Million cell updates/sec

Title: US-10-092-390-1

Perfect score: 3423
Sequence: 1 atggtatttttttgtaactc.....gcagcagcagcagtgatga 3423

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_luv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_frod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1334.2	39.0	3162	11	BC029999	BC029999 Homo sapi
2	456.6	13.3	755	12	BC628819	BC628819 602751595
3	433	12.6	565	14	BM719978	BM719978 UI-E-EJ0-
4	432	12.6	598	14	BM676825	BM676825 UI-E-EJ0-
5	379	11.1	538	10	BM762574	BM762574 BM762574
6	370	10.8	715	9	AI958909	AI958909 fd05g03.y

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REMARK	COMMENT
BC029999	BC029999	Homo sapiens, clone IMAGE:4156083, mRNA.	BC029999	BC029999.1	GI:20455873	Homo sapiens.	1	Strausberg, R.	Direct Submission	Submitted (06-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-rt@mail.nih.gov Tissue Procurement: David N. Louis, M.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu	

ALIGNMENTS

7	367.6	10.7	438	13	BM484594	BM484594 538509 MA
8	366.8	10.7	779	12	BF529240	BF529240 602041695
9	359.6	10.5	641	10	BB623063	BB623063 BB623063
10	354.6	10.4	488	14	HI5471	HI5471 ym29502.r1
11	328.2	9.6	584	12	BF686873	BF686873 602102830
12	309.4	9.0	1123	13	BM563529	BM563529 AGENCOURT
13	302	8.8	559	10	AM658138	AM658138 93921 MAR
14	296.2	8.7	937	12	BF180097	BF180097 601806458
15	294.4	8.6	852	9	AL040177	AL040177 DKFZP434F
16	293.2	8.6	401	10	BE664785	BE664785 152639 MA
17	282.2	8.2	463	10	AM445546	AM445546 81787 MAR
18	271.2	7.9	477	12	BF442243	BF442243 258883 MA
19	260	7.6	501	10	AM478869	AM478869 22097 MAR
20	253.6	7.4	634	10	BB650216	BB650216 BB650216
21	244.2	7.1	265	14	Z43452	Z43452 HSC1CER041 n
22	240.6	7.0	675	10	BB352543	BB352543 BB352543
23	214.8	6.3	732	12	BG702195	BG702195 602684158
24	214	6.3	784	13	BI829828	BI829828 603079907
25	213.4	6.2	345	14	RI4925	RI4925 yf494602.r1
26	210	6.1	543	12	BF079099	BF079099 229629 MA
27	199.2	5.8	668	13	BM427205	BM427205 p9f2n.pk0
28	197	5.8	723	17	AQ239321	AQ239321 RPI111-68
29	189.4	5.5	643	10	BB866292	BB866292 BB866292
30	182.8	5.3	627	13	BI445567	BI445567 dae82d04.
31	177.4	5.2	332	13	BI653297	BI653297 603301011
32	176.8	5.2	939	14	BO919892	BO919892 AGENCOURT
33	172.4	5.0	876	12	BF347336	BF347336 602020612
34	169.8	5.0	890	13	BI697039	BI697039 603348907
35	163.4	4.8	858	13	BG920104	BG920104 602821285
36	158.6	4.6	638	10	BB656858	BB656858 BB656858
37	157.6	4.6	741	14	BM715446	BM715446 UI-E-EJ0-
38	151	4.4	645	13	BM652416	BM652416 170006873
39	150.4	4.4	920	14	BO919101	BO919101 AGENCOURT
40	144.8	4.2	448	10	BB647398	BB647398 BB647398
41	141.2	4.1	522	14	BO554245	BO554245 H4026H12-
42	140.8	4.1	1029	17	AL276426	AL276426 Tetradon
43	139.2	4.1	434	12	BG383545	BG383545 301686 MA
44	139.2	4.1	652	12	BG749427	BG749427 602077339
45	137.6	4.0	872	12	BG489428	BG489428 602517788

Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAX Plate: 52 Row: a Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA g1: 14192940
This clone has the following problem: no 5' EST match.

FEATURES

source

Location/Qualifiers

1. 3162
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4156083"
/tissue="type="Brain, anaplastic oligodendroglioma with
1p/19q loss"
/clone.lib="NCI CGAP_Brn67"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 592 a 1021 c 938 g 611 t
RIGIN

Query Match 39.0%; Score 1334.2; DB 11; Length 3162;
Best Local Similarity 69.4%; Pred. NO. 0;
Matches 1845; Conservative 0; Mismatches 808; Indels 6; Gaps 2;

QY 74 CTCTGAATCTTGAAGACCTTATGTGTAGCACTGGGAAAGCTACTGACTGTGC 133
DB 229 CCTGAACCCCGAGGACCCGAAAGTGTGACGACCTGGAGAGCTATGCTGTGCC 288
QY 134 AAGAGCTATACCCACATCCCTTTGATCAAAATTACTACAGAGCTGACACTTCTAA 193
DB 289 AGGAATCTATGACACACCCCTTCAGATCTATTACACAGATGACAGACATCTCA 348
QY 194 ACTGTTAAATGACGGGCGACAGATGACATGAGGACGACCTATGACATGGGAGA 253
DB 349 ACTGTTAAAGTGCACAGGACCGGATCAGTTATTAAGCGCGATGAGAGGCTTC 408
QY 254 AGACTATATAGGCGCAAGTCTCAGTGTGTCTGATTTTAAAGCGGGAATGT 313
DB 409 GGACCATGTACGGGGGAGGTCCAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCT 468
QY 314 GTGTCCCACTGTCTGTAAATGTGTCAATGCTGCTGTATGCTGCTGCTGCTGCT 373
DB 469 GCATACCCCTGTGAGGAGAGTGTGACAGGCGCGCTGCTGCTGCTGCTGCTGCT 528
QY 374 AGTGTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 433
DB 529 ACTGTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 588
QY 434 GTCCCACTGACACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 493
DB 589 GGCCCACTGACACAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 648
QY 494 GGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 553
DB 649 GCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 708
QY 554 GCACCTATGTAAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 613
DB 709 GCACCTATGTAAGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
QY 614 ACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 673
DB 769 CCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 828
QY 674 GTCTCTCTGTAAACATGCTGACAGTGTGACAGAGATGCTGCTGCTGCTGCTGCTGCT 733
DB 829 GCGCTCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 888
QY 734 TGTGTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 793

DB 889 CTTGCAACACATCATCTGCGAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 948
QY 794 GTACACCTTGGCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 853
DB 949 CCGACCCCTGCGCCACACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1008
QY 854 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913
DB 1009 ATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1068
QY 914 AACGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 973
DB 1069 ACAGTGTGCAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1128
QY 974 AGTGTGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1033
DB 1129 ACTGCGCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1188
QY 1034 TTGCTGTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1093
DB 1189 ACAAGGCGCCACAGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1248
QY 1094 ACAAGGCGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1153
DB 1249 CCTGTGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1308
QY 1154 GTGCTGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1213
DB 1309 GTACTGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1368
QY 1214 ACGGGAAGCTTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1273
DB 1369 ATGGGATGGTGTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1428
QY 1274 CTGGAATGTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1333
DB 1429 CTGGGGGTGCACTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1488
QY 1334 TGGGAACCTATGGAATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1393
DB 1489 CAGGAGCTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1548
QY 1394 CTCTGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1453
DB 1549 CCCCAGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1608
QY 1454 GATGTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1513
DB 1609 CATGTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1668
QY 1514 GAGCTGTGCAACCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1573
DB 1669 CAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1728
QY 1574 GCGAATCTTCCCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1633
DB 1729 GTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1788
QY 1634 GCGACGAGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1693
DB 1789 GCGATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1848
QY 1694 GTGTGTCACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1753
DB 1849 GCATGTGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1908
QY 1754 GCTACTGTAAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1813
DB 1909 GCAGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1968
QY 1814 GCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1873
DB 1969 GCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2028

1874 GGCAGACATGCCCCACAGTGGCTTACAGAGAGGGGCCCCCTGCACCAATCACCAGCCCTCT 1933
 2029 CCAGGACATGCCCCCTCTCGTGCACAGAGAGCCCTGCGACCAATCACCAGCCCTCT 2088
 1934 GTGACTGCTTGGCTGGCTTGCACAGGCGCCCTGCAATGAATGTGTCCAGTGGAGAT 1993
 2089 GTGAGTGGCTCCAGAGATTTCTGTGAGAGCTCTGTGACCAACCAATGTGTGTGTGAGAT 2148
 1994 TTGGGAAAACTGTGAGAGAAATTTGTACTGTGACCAACAGAGGAACTGTAAACCCAT 2053
 2149 TTGGGAGAGAGCTGTGCGCCAGCTCTGTCTGTGCAACAGAGGAACTGTGACCCAT 2208
 2054 ACAGATCTTGTGAGTGTACCCCGGTGTGATGTGAGAGTGTGACCTGTCTCAACATGTCC 2113
 2209 ATGGCTCTGTGCGAGCTTCTGTGAGATTTGTGAGAGAGAGCTGTGACAGGCTTGTCC 2268
 2114 CTGCCACTGTGGGGCCCAAACTGCATCAGACGCTGCACATGATGAGAGCTTGTCTGCA 2173
 2269 CGGGGTCTGGGGGGCCCCGCTGCTGTCCAGCATGACAGCTGCACAGAGGGGCGAGCTGCA 2328
 2174 GGGCCTAGCATGGGAGATTAATGCACTCTGGCTGTGACAGAGGCTGTACTGCACTAGA 2233
 2329 GGGCGAGAGAGGGGGCTGCTCCAGTGCACCCCTGCTGAGAGCTGTCTGTGACAGAGC 2388
 2234 GATGCTCTGTAGGGTTTATGAAAAAATGTGCACTGATGATGCAATGTCAAAACGAG 2293
 2389 GTTGCCAGACATTTTGTGGAAAGAGCTGTGGGGCGGTATGCTCCAGTGTCAAGAAATGG 2448
 2294 CTGACTGTGACCAATTTCTGGGAGTGTGCTGCTGCGCACTGATGATGATGATGATGAT 2353
 2449 CAGGCTGTACCAATCAATGTGCAAGTGTGCACTGCTGCGCACTGATGATGATGATGAT 2508
 2354 GTGAGCAAGAGTGGCTTGTGAGAAATGTGCTGATGCTGTGCTGCTGCTGCTGCTGCT 2413
 2509 GTGAGCAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2568
 2414 TGAACAACTGCACAGTGCACATCAGTGCAGAGGAGTGTGCTGCTGCTGCTGCTGCTGCT 2473
 2569 TGAACAACTGCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2628
 2474 GAGCAGATGTATCAAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2533
 2629 GATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2685
 2534 CAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2593
 2686 TCAGCCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2745
 2594 TTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2653
 2746 TGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2805
 2654 GAAAGAAATCAAGCATG--CCAGCAGTATCACTACACCCCTGCTGATGAGGCTGCTCAAT 2710
 2806 ACAAGGCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2865
 2711 CAGATTATCACTTTCAGG 2729
 2866 CCGACTACTCCTCTCAGG 2884

RESULT 2
 LOCUS BG828819 755 bp mRNA linear EST 22-MAY-2001
 DEFINITION 602751395F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4904255 5',
 mRNA sequence.
 ACCESSION BG828819
 VERSION BG828819.1 GI:14176418
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 755)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L16M1803 row: m column: 24
 High quality sequence stop: 707.
 Location/Qualifiers
 1..755
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="4904255"
 /clone_lib="NIH_MGC_17"
 /tissue_type="Thadomyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: muscle; Vector: pORF; Site:1: EcoRI;
 Site:2: XhoI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 183 a 175 c 197 g 200 t
 ORIGIN

Query Match 13.3%; Score 456.6; DB 12; Length 755;
 Best Local Similarity 98.8%; Pred. No. 1.3e-122;
 Matches 481; Conservative 0; Mismatches 4; Indels 2; Gaps 2;
 1 ATGGTTATTTCTTGTGAACTGATGCTGAGCTTATTTGTTATTTGTTATGCTGATG 60
 268 ATGGTTATTTCTTGTGAACTGATGCTGAGCTTATTTGTTATTTGTTATGCTGATG 327
 61 GGGACAGCATCACTGTAATCTGTAAGACCTTAATGTGTAGCACATGGGAAAGCTAC 120
 328 GGGACAGCATCACTGTAATCTGTAAGACCTTAATGTGTAGCACATGGGAAAGCTAC 387
 121 TCAGTACTGTGCAAGAGTATACCCACATCCCTTTGATCAATTTACTACAGAGCTGC 180
 388 TCAGTACTGTGCAAGAGTATACCCACATCCCTTTGATCAATTTACTACAGAGCTGC 447
 181 ACTGACATTTCTAATCTGTTTAAATGACGCGGACAGAGTACGATGAGACAGCTAT 240
 448 ACTGACATTTCTAATCTGTTTAAATGACGCGGACAGAGTACGATGAGACAGCTAT 507
 241 CGACATGGGGAAGAGTATGATGAGGCGGACAGTCAAGTGTGCTGCTGATTTATGAA 300
 508 CGACATGGGGAAGAGTATGATGAGGCGGACAGTCAAGTGTGCTGCTGATTTATGAA 567
 301 AGCGGGGAATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 359
 568 AGCGGGGAATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
 360 TCCTAAACACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 419
 628 TCCTAAACACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687
 420 TGGTATCACTGCGGTGCCCATGACAGCGCGGTG--CCAGTCAAAATGGGGCTGTGT 478
 688 TGGTATCACTGCGGTGCCCATGACAGCGCGGTG--CCAGTCAAAATGGGGCTGTGT 747
 479 GCMAACC 485

Db 748 GCCACCC 754

RESULT 3
BM719978LOCUS 565 bp mRNA linear EST 01-MAR-2002
DEFINITION UI-E-EJ0-ahu-1-16-0-UI.1 UI-E-EJ0 Homo sapiens cDNA cloneACCESSION BM719978
VERSION BM719978.1 GI:19038910KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 565)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).

Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1..565

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahu-1-16-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"

/dev_stage="fetal and adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/note="Organ: eye; Vector: pRT73-Pac (Pharmacia) with a
modified polylinker; Site.1: EcoR I; Site.2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pRT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, ACATACAGCA
; lens, GCATTAGCGA; eye anterior segment, AATGCCGAT;
optic nerve, CCATTAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 179 a 122 c 139 g 125 t

ORIGIN

Query Match 12.6%; Score 433; DB 14; Length 565;
Best Local Similarity 100.0%; Pred. No. 9.5e-116;

Matches 433; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2800 ACCGAGTGTGCACATCCCTCAGTCACACAGGAGATGATGTCACGAGTCA 2859

Db 1 ACCGAGTGTGCACATCCCTCAGTCACACAGGAGATGATGTCACGAGTCA 60

QY 2860 AAAAAATCAACTGTTTGTGAACTTAAAAATGTGAACCTGGGAAGAGAGCCCTGTG 2919
 Db 61 AAAAAATCAACTGTTTGTGAACTTAAAAATGTGAACCTGGGAAGAGAGCCCTGTG 120
 QY 2920 GGGAGCTGACATGGGACATATGCCGGCTGACTGGAAAAATGGCGCTACCTCAAGACTC 2979
 Db 121 GGGAGCTGACATGGGACATATGCCGGCTGACTGGAAAAATGGCGCTACCTCAAGACTC 180
 QY 2980 GGTCTTTTGGACTTGCACAGAGCTATATGGGAAATCTTAAAAAGACTGGGAAAGAT 3039
 Db 181 GGTCTTTTGGACTTGCACAGAGCTATATGGGAAATCTTAAAAAGACTGGGAAAGAT 240
 QY 3040 TCTGAATATATCAATCACTGCTCCCTAAGCAGTTCTGAGAACCAATATGCCATATT 3099
 Db 241 TCTGAATATATCAATCACTGCTCCCTAAGCAGTTCTGAGAACCAATATGCCATATT 300
 QY 3100 AAGACCCACCTGACTTATCCCAAGAAAGCTCAAGAGTGGTTATGGGAGTAAATTCG 3159
 Db 301 AAGACCCACCTGACTTATCCCAAGAAAGCTCAAGAGTGGTTATGGGAGTAAATTCG 360
 QY 3160 CCGCAGAGAGAGATTCCTCATATGACAGATCAATCACTCACTTCAAGCAGAGAT 3219
 Db 361 CCGCAGAGAGAGATTCCTCATATGACAGATCAATCACTCACTTCAAGCAGAGAT 420
 QY 3220 GTCATGAGATTG 3232
 Db 421 GTCATGAGATTG 433

RESULT 4
BM676825/c

LOCUS 598 bp mRNA linear EST 27-FEB-2002

DEFINITION UI-E-EJ0-ahu-1-16-0-UI.s2 UI-E-EJ0 Homo sapiens cDNA clone

ACCESSION UI-E-EJ0-ahu-1-16-0-UI 3', mRNA sequence.
BM676825

VERSION BM676825.1 GI:18986721

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 598)

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).

Seq primer: M13 forward

FEATURES

source

Location/Qualifiers

1..598

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="UI-E-EJ0-ahu-1-16-0-UI"

/clone_lib="UI-E-EJ0"

/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"

/dev_stage="fetal and adult"

/lab host="DH10B (Life Technologies) (T1 phage resistant)"
 /note="Organ: eye; Vector: p773-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 01-E-EJ0 is a subcloned cDNA library constructed
 according to Bonaldo, Lennon and Soares, Genome Research,
 6:791-806, 1996. First strand cDNA synthesis was primed
 with an oligo-dT primer containing a Not I site. Double
 stranded cDNA was ligated to an EcoR I adaptor, digested
 with Not I, and cloned directionally into p773-Pac
 vector. The oligonucleotide used to prime the synthesis of
 first-strand cDNA contains a library tag sequence that is
 located between the Not I site and the (dT)₁₈ tail. The
 sequence tags for this library are: fetal eyes, AGATCAGAGA
 : lens, CGATTAGCGA: eye anterior segment, AATGCCGCGAT;
 optic nerve, CCATTAAGTG; retina, CCGCG; Retina foveal and
 Macular, GTCC; RPE and Choroid, ACCCA. This library was
 created for the program, Gene Discovery in the Visual
 System, supported by National Eye Institute (NEI).
 TAG, LIB-UI-E-EJ0
 TAG, TISSUE-human fetal eyes
 TAG_SEQ=AGATCAGAGA"
 BASE COUNT 130 a 146 c 123 g 199 t
 ORIGIN

Query Match 12.6%; Score 432; DB 14; Length 598;
 Best Local Similarity 98.9%; Pred. No. 1.9e-115;
 Matches 435; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2793 CAGCTCACCAGTGTGCACATCCCTCAGTCACACAGAGGACAGATGACTGTAC 2852
 DB 596 CAGGAGAGCCAGTGTGCACATCCCTCAGTCACACAGAGGACAGATGACTGTAC 537
 OY 2853 GAAGTCAAAAACATCACTGTTTGAATCTTAAATGTGAACCTGGGAAGAGAG 2912
 DB 536 GAAGTCAAAAACATCACTGTTTGAATCTTAAATGTGAACCTGGGAAGAGAG 477
 OY 2913 CCTGTGGGGGAGTGCATGCGATGCGGCTGAGTGAACATGCGGCTACTCTAA 2972
 DB 476 CCTGTGGGGGAGTGCATGCGATGCGGCTGAGTGAACATGCGGCTACTCTAA 417
 OY 2973 CGAGCTCGGTGCTTTGGACTTGACAGAGCTATATGGAAGAAATCCTTAAAGACCTGG 3032
 DB 416 CGAGCTCGGTGCTTTGGACTTGACAGAGCTATATGGAAGAAATCCTTAAAGACCTGG 357
 OY 3033 AAGATTTGATATATCAATCAAGTACGTCCTCAACAGCTGTGAGACCCATATGC 3092
 DB 356 AAGATTTGATATATCAATCAAGTACGTCCTCAACAGCTGTGAGACCCATATGC 297
 Y 3093 CACTATTAAAGACCCACTGTACTTATCCGAAAGCTCAGAGTGTGTATGTGAGAT 3152
 DB 296 CACTATTAAAGACCCACTGTACTTATCCGAAAGCTCAGAGTGTGTATGTGAGAT 237
 OY 3153 GAATTCGCGGACGAGAGGATTTCCCAATATGACAGATCAATTAATCACTTACGCCAA 3212
 DB 236 GAATTCGCGGACGAGAGGATTTCCCAATATGACAGATCAATTAATCACTTACGCCAA 177
 OY 3213 CAGGAATGCTATGAAGTTG 3232
 DB 176 CAGGAATGCTATGAAGTTG 157

RESULT 5 538 bp mRNA linear EST 17-OCT-2001
 BB762574
 LOCUS BB762574
 DEFINITION BB762574 RIKEN full-length enriched, B16 F10Y cells mus musculus
 accession BB762574
 version BB762574.1 GI:16207666
 keywords EST.
 source house mouse.
 organism Mus musculus
 Mammalia; Eutheria; Rodentia; Sclerognathii; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 538)
 AUTHORS Akimura, T., Akakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T., Imtani, K., Ishii,
 Y., Ito, M., Kawai, D., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
 Nakamura, M., Nishi, K., Nomura, K., Nunasaki, R., Okazaki, Y., Okido, T.,
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
 Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,
 A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toyu, T.,
 Wataniki, A., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001)
 TITLE Unpublished (2001)
 JOURNAL Contract: Yoshitake Hayashizaki
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel.: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Ito,
 M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new
 genes. Genome Res. 10 (10), 1617-1630 (2000)
 Wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
 Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura,
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A., and
 Hayashizaki, Y.
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichannel sequencer. Genome Res.
 10 (11), 1757-1771 (2000)
 Konno, H., Fukunishi, Y., Shibata, K., Ito, M., Carninci, P., Sugahara,
 Y., and Hayashizaki, Y.
 Computer-based methods for the mouse full-length cDNA
 encyclopedia: real-time sequence clustering for construction of a
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for
 further details.
 e mouse tissues.

FEATURES

source location/Qualifiers
 1..538
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="G370013A07"
 /clone_11b="RIKEN full-length enriched, B16 F10Y cells"
 /cell_type="B16 F10Y cells"
 /note="pooled tissues: (tissue_type=cerebellum,
 dev_stage=16 days neonate, sex=mixed),
 (tissue_type=cerebellum, dev_stage=0 day neonate,
 sex=mixed), (tissue_type=hippocampus, dev_stage=adult,
 sex=male), (tissue_type=whole body, dev_stage=9 days
 embryo, sex=mixed), (tissue_type=lung, dev_stage=13 days
 embryo, sex=mixed)"
 BASE COUNT 161 a 142 c 131 g 104 t
 ORIGIN

Query Match 11.1%; Score 379; DB 10; Length 538;
 Best Local Similarity 86.3%; Pred. No. 7.6e-100;
 Matches 465; Conservative 0; Mismatches 70; Indels 4; Gaps 4;

OY 2652 GGAAGAAGATACAGATCCGACGACTTACACACCCCTGTATGAGGAGGCTCAATGC 2711
 DB 1 GAGGAAGAATACAGATCCGACGACTTACACACCCCTGTATGAGGAGGCTCAATGC 60
 OY 2712 AGATTATACACTTACAGAACCTTCCACAGCAATGCTGAAGAGCTAATAGCACATTA 2771
 DB 61 AAGCTATACACTGCAAGAAACCTGCTACAGCAATGCT-6AATAGCAAGACACATTA 119
 OY 2772 CTTACACATCCAGTTCACACACGCTCAGCTGTCACATCCCTCAGCTCAACAA 2831

Db 662 TCATGTGTGTTCATGAGAGCGCATCATGCTCTGATGGCA 707

RESULT 7
BM484594 438 bp mRNA linear EST 05-FEB-2002
LOCUS 536509 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BM484594
ACCESSION BM484594.1 GI:18534922
VERSION EST.
KEYWORDS
SOURCE
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
REFERENCE
1 (bases 1 to 438)
Fahnensterg, S.C., Fickling, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
and Keeler, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
CONTACT: Smith TP
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemall.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross-match with the -m1nscore 18
and -m1smatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCATCAGCAGC
Plate: 4 row: B column: 20
Seq primer: ATTTAGTGACATATGAC.
FEATURES
source
1. 438
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; site.1: NotI; site.2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 133 a 105 c 101 g 99 t
ORIGIN
Query Match 10.7%; Score 367.6; DB 13; Length 438;
Best Local Similarity 90.0%; Pred. No. 1,5e-96;
Matches 394; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Db 2862 AAACATCACTGTTTGTAATCTTAAATGTGAACCTGGAGAGAGCCCTGTGGG 2921
1 AAATATACAGTGTGTTGTAATCTGAAAATGTGAATCCTGGAAGAGCCCATGGC 60

Db 2922 GGACGCACTGGGACATTCGCGGTGACGGAACATGGCGCTTCAACGAGCTCG 2981
61 GGACATATAGGGGAGCGCTGCGAGACTGGAACATGGCGCTTCAACGAGCTTGG 120

Db 2982 TGCCTTTGGACTTGACAGAAAGCTATATGGAATTCCTTAAAGACCTGGGAAAGATTG 3041
121 TGCCTTTGGACTTGACAGAGGTTATATGGAATTCCTTGAAGACCTGGGAAAGATTG 180

Db 3042 TGAATATATTAAGTAAGTCTGCTTAAGCAAGTCTGGAACCATATGCACTATTAA 3101
181 TGAATATATTAAGTAAGTCTGCTTAAGCAAGTCTGGAACCATATGCACTATTAA 240

Db 3102 AAACCCACTGTACTTATCCCAAAAGCTCAAGAGTGTGTTATGTGGAGATGAATCGCC 3161
241 AAACCCACTGTACTTATCCCAAAAGCTCAAGAGTGTGTTATGTGGAGATGAATCGCC 300

Db 3162 GGCACAGAGAGATTCCTCATATGACAGATCAATCACTTACAGCCACAGGAATGT 3221
301 GGCCTGGAGAGAGCTCCCATATGACAGAGATCAATCACTTACAGCCACAGGAATGT 360

Db 3222 CATAGAGTTGAACCTACAGTGTGTGTCACAGAGATTTACCAATATGCGCTCT 3281
361 CATAGAGTTGAACCTACAGTGTGTGTCACAGAGATTTACCAATATGCGCTCT 420

Db 3282 CTCGCCAGATCCATATGA 3299
421 CACTCAGATCCCTATGA 438

RESULT 8
BF529240 779 bp mRNA linear EST 11-DEC-2000
LOCUS 602041695F1 NCI_CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4179646
DEFINITION 5', mRNA sequence.
ACCESSION BF529240
VERSION BF529240.1 GI:11616603
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 779)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
CONTACT: Robert Strusberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM9489 row: m column: 23
High quality sequence stop: 737.
FEATURES
source
1. 779
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4179646"
/clone_lib="NCI_CGAP_Brn67"
/tissue_type="anaplastic oligodendroglioma with 1p/19q
loss"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: brain; Vector: pCMV-SPORT6; site.1: NotI;
site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 132 a 245 c 239 g 163 t
ORIGIN
Query Match 10.7%; Score 366.8; DB 12; Length 779;
Best Local Similarity 70.2%; Pred. No. 3.8e-96;
Matches 550; Conservative 0; Mismatches 22; Indels 7; Gaps 4;

Db 1159 TGCAGCGGGGTGTCAGAGCTACTGATATGAGATGTTCTCTGATTTACGGG 1218
1 TGCAGCGAGGCTGTGTGTGTCACACATCATGATCTGCTGTTGTGCTACTATGGC 60

Db 1219 GAAGTTGGCAGCAGATCTGAGCTGCCAAAATGGGGAGAGCTGACAGTGTGCTGA 1278
61 GATGCTGGCAGCTGCTGCTGTCACATGATGCGCGAGCTGCAACAGATCACTGGG 120

Db 1279 AAGTGCACCTGTGCCCCAGAGATCAAGAATGATGCTCTACACCCATGCCCTTGGA 1338
121 GGTCTACTGTGCTCCGGCTTCATGGAGAGGCTGTGCGTTCTCTGACAGAGG 180

Db 1339 ACCATGGAGTAACTGTTCTGCTGCTGTGAGCTGTAAAAATGATGATGCTGCTCCT 1398

Db 121 CACCGGATTCAGGTATTAAGACCCGGATATAGCGCGGCCCTCCGGACCATGTACCGGGGAG 180

Qy 274 TCTCAGTGTGTCTCTGATTTTATGAAGCGGGGAAATGTGTGTCGCCACTGTGCTGAT 333

Db 181 TCCCATGCTCGCCCGCTACTATGAGAACGGAGACTTCTGCATTCCTCTGTGTACGAG 240

Qy 334 AAATGTGCCATNGTGTGCTGTATTTCCTCCAAACACCTGTCAAGTGTGAGCTGTGGGGA 393

Db 241 GAGTGCATGACACGGCGCGCTGTGTCTCTCCGATACCTTCCCACTGTGAGCTGTGAAGGGA 300

Qy 394 GGGACCAACTGCTGCAGTGTGCGCATGTGATCAGTGTGGGTCCTGACGACACGAGCGG 453

Db 301 GGGCCTGACTGTGTCCAGGGGCTGTGACAGCGAGCACTGGGGTCCCACTGCACGACCGG 360

Qy 454 TGCCAGTGCAAAATAATGGGGCTCTGTGCAACCCCATCACCGGGGCTTGGCACTGTGCTCG 513

Db 361 TGTGAGTGTCAAGACGGGGCCTGTGCAACCTATCACCGGGGCTGTGTGGGCCCCG 420

Qy 514 GGCCTTCGGGGCTGGCGCTGCGAGGACCGCTGTGAGCAGGGCACCCTATGTATACGACTGT 573

Db 421 GGCCTTCGAGGCTGGCGCTGTGAGGAACCTGTGCGCTCTGTGTATCTCAAGGGCAAGGGCTGC 480

Qy 574 CATCAGATATCCAGTGTCCAGATATGAGAGCCACTGTGCACCAAGTACAGGGGGAAATGCCGC 633

Db 481 CAGTGTCTGTTCAGATGTGCACCATGTGAGCGGCCCAAGCTGTGACCCGGCAGCTGCGGAGTGCCTC 540

Qy 634 TGCCACACCGAATACCCGGAACCTTCTGTGAGAGATCTTTGTCTCTCTGTGTAAACATGGT 693

Db 541 TGCGGTCTCTGGCTACACAGGGCGTTACTGTGAGGAGCTGTGCACCCCTGGGAGCATGGA 600

Qy 694 CCACAGTGTGAGCAGAGATGCCCTGTGCAAAATGAGG 731

Db 601 GCTACTGTGAGCTGCGCTGCGCCCTGCGCCAGATGAGG 638

```

/db_xref="taxon:9606"
/clone_image:49332"
/clone_1lb="Soares infant brain 1N1B"
/sex="female"
/dey_stage="73 days post nata)"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: lafmid BA; Site:1: Not
I, Site_2: Hind III; 1st strand cDNA was primed with a Not
I, -olligo(dN) primer 15'
AAGCGAGAGAAATTCGCGCCGACGAAATTTTTTTTTTTT 3'};
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the lafmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fátima Bonaldo."

```

Query Match	10.4%	Score 354.6	DB 14	Length 488
Best Local Similarity	93.8%	Pred. No. 1.1e-92		
Matches 425, Conservative	0	Mismatches 19	Indels 9	Gaps 5

Db 414 CAGTAACAGAGCTGTACCTCCAGCCAGGCTGTGTGTCACCACTGCAATGATCCT 473
 Oy 1199 GTTCTCTGATGATCTACAGGGAGAGCTGCGACAGATGTGACGTGCCAAATGGGCGG 1258
 Db 474 GCCCTGTGCTACTAGTGCATGGCTGACCTGCTTGCACCTGTGCAAGATGGGCGG 533
 Oy 1259 ACTGTGACAGTGTGACATGGAAGTGCACCTGTGCCCCAGAGATTAAGAGATGACTGCT 1318
 Db 534 ACTGCGACAGCATCACTGGGGGCTGACATGTGCTCCGGCTTCAATGGAGAGCTGTG 593
 Oy 1319 CTACCCCATGCCCTCTGGGAACCTATGGGATACTGTCTCTGCTGTGGCTGTAAAA 1378
 Db 594 CCGTTTCTGTGCTACAGAGAGCTATGGCCCACTGCTGCTCAATCTGTACTGTAAACA 653
 Oy 1379 ATGATGAGATGCTGCTCTCTGTGGAAGGT-CTTGTACTTGTGAAGCA--GGCTGGCAG 1435
 Db 654 ATGTGGGACCTGCTCCAGAGATGCTCTCTGTACTGTGAAGAAAGGGTGGCAGNG 713
 Oy 1436 GGGTGAAGTGTCTCATAGATGTCCCACTGACATGGGAGCTTGTGGCTTAATCAAT 1495
 Db 714 CCTTGAGTGCACCTGCGCATGTCCAGTGGGAGCTGGGGGCTGAATGCAACGAGAGCT 773
 Oy 1496 GCCAATGCTCTCAAGGGGGAGCTGCAACCTGAGAGGGAGCTGCACAGTGTGACCTG 1555
 Db 774 GCACCTGTGCTCAATGGGGAGCTGCAACCTGATACGCTGCTCCGGCTCCGAGCTCCGG 833
 Oy 1556 GATGCGGGGGGAGAAATGCGAATCTCCCTG-CCAGAGATGCGAGCGAGGCTGACT 1612
 Db 834 GCTGCTGGAAACCCCTGTGATTTGCTCTGCTCCCGGAGGAGGACCTTTGGGCTGACT 891

RESULT 13
 AM658138 559 bp mRNA linear EST 25-APR-2001
 LOCUS
 DEFINITION 93821 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION AM658138
 VERSION AM658138.1 GI:7423964
 KEYWORDS EST.

SOURCE

ORGANISM
 Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.

REFERENCE

AUTHORS
 Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
 Casas,E., Wray,J.E., White,J., Cho,J., Fahnenkrug,S.C., Bennett,
 G.L., Heaton,M.P., Jaeger,D.W.W., Kohrer,G.A., Chitko-McKown,C.G.,
 Pettea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and
 Keeler,J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle

JOURNAL

MEDLINE
 Genome Res. 11 (4), 626-630 (2001)

COMMENT

Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smithemall.marc.usda.gov
 Single pass sequencing. Bases called and alt-trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -minscore 18
 and -minmatch 12 options.

PCR

Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCCTCAGCTACGACG
 Plate: 60 row: 1 column: 12
 Seq primer: ATTAGGTGACACTATAG.

FEATURES

source

1..559
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_id="MARC 1BOV"
 /tissue_type="pooled"

/lab_host="DH10B"
 /note="Vector: pCMV SPOR6; Site.1: NotI; Site.2: SalI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."

BASE COUNT
 ORIGIN

86 a 182 c 181 g 110 t

Query Match 8.8% Score 302; DB 10; Length 559;
 Best Local Similarity 71.3%; Pred. No. 3.9e-77;
 Matches 398; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Oy 1226 GCCACAGATCTGACAGCTGCCAAATGGGGGAGAGTGTGACGTGTGAGAAATGCA 1285
 Db 2 GCCAGCTGCTGTGACACTGTCTGCAAGACGGTCCGAGATGCCAGCATCACTGGAGACTGA 61
 Oy 1286 CCTGTGCCCGCAGATTCAGAAAGATTAATGACTGCTCTTACCCCATGCCCCCTGGACCTATG 1345
 Db 62 CTTGTGCCCCAGAGCTCATGGGGGAGAGTCTGTGCAAGCTCTCTGCGGTGGGGAGCTATG 121
 Oy 1346 GGAATAACTGTCTCTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1405
 Db 122 GCTCCAGCTGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 181
 Oy 1406 GCTCTGTACTTGTCAAGGAGGCTGTGCAAGGGGGTGGAGTGTCTCATAGATGTCCAGATG 1465
 Db 182 GCTCTGTGACCTGCAAGGAGGGGTGGAGGGCTGTGACAGTGCACCTGCGGTGCCAGATG 241
 Oy 1466 GCACATGGGGCTTTGGCTGTAACTTAATGACATGCTGCTCAAGGGGGAGGAGCTGCACAA 1525
 Db 242 GGACGTGGGGCTGTGAATGCAATGAGAGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 301
 Oy 1526 CCTGTGAGGAGCTGTGACAGTGTGACCTGTGATGGCGGGGAGAAATGCAACTTCCCT 1585
 Db 302 CCACAGATGATGATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 361
 Oy 1586 GCCAGATGGACAGTACGGGCTGAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1645
 Db 362 GCCCGATGGACATTTGGGCTGAACTGTGAGTGAACGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 421
 Oy 1646 GCTGCCACCTTACACAGGGGCTTTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1705
 Db 422 GTTGGACCCCTGTGACAGCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481
 Oy 1706 ACAGCTGTGCTGCTGAGGAGCTGCGGGGCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1765
 Db 482 ACAGCAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 541
 Oy 1766 ATGGGGCTTCATGCTTCC 1783
 Db 542 ATGGCGTTCTCTGCTGCC 559

RESULT 14

JOURNAL

MEDLINE
 DEFINITION 937 bp mRNA linear EST 31-OCT-2000
 ACCESSION BF180097
 VERSION BF180097.1 GI:11056239
 KEYWORDS EST.

SOURCE

ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS
 NIH-MGC http://mgc.ncl.nih.gov/
 TITLE
 NATIONAL Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Contact: Robert Strausberg, Ph.D.

COMMENT

Email: cgabs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)


```
OY 1454 GATGTCCAGTGGACATGGGGCTTGGCTGTAACATAACATGCCAGTGCCTCAACGGGG 1513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 541 CTTGCCCAACCCGGAACCTGGGGCTTCAGTTCATGCCAGCTGCCAGTGTGCCCATGAGG 600
OY 1514 GAGCCTGCACACCCCTGGACGGGACCTGCACGTGTGCACCTGATGGCGGGGAGAAAT 1573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 601 CAGTCTGCAGGCCCCCAAACTGAGACCTGTACTGCACCCCTGGGTGGCATGGGCCCACT 660
OY 1574 GCGAACTTCCCTGCCAGAGTGGCACGTACGGGCTGAACTGTGTGAGCGCTGCGACTGCA 1633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 GCCAGCTGCCCTGTCCGAGGGGCGAGTTGGAGAAGGTGTGCCAGTGGCTGTGACTGTG 720
OY 1634 GCCACGAGATGGCTGCCACCCCTACACAGGGCCATGTGCCGCTGCCGGGATG-GTCA 1692
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 ACCACTCTGATGAGCTGTGACCCCTGTTCATGAGACGCTGTCACTGCCAGGCTGGGATG 780
OY 1693 GGTGTCCACTGTGTACAGCGTGTGTGTGAGGAGACGTGGGGCCCCAACTGCTCCCTGCC 1752
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 781 GGTGCCCGCTGGCCACCTGTCTCTGCTGAGGGCTTATGGGAGTCAACTGTAGCAACACC 840
OY 1753 TGTACTGTATAA 1764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 841 TGCACCTGCAGA 852
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: May 9, 2003, 11:39:44
Job time : 4608 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 9, 2003, 07:45:02 ; Search time 403 Seconds
(without alignments)
10555.283 Million cell updates/sec

Title: US-10-092-390-1
Perfect score: 3423
Sequence: 1 atggtattcttctgaactc.....gcagcagcagcagtgatga 3423

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

otal number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_MA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCr_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCrUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3423	100.0	3423	9	US-10-092-390-1
2	1760	51.4	1761	9	US-10-092-390-3
3	1425.8	41.7	1448	12	US-10-105-929-9
4	608.4	17.8	3567	9	US-09-796-753-99
5	608.4	17.8	3567	9	US-09-796-753-123
6	528.6	15.4	5036	9	US-09-796-753-113
7	524.6	15.3	5197	10	US-09-833-381-1910
8	424.8	12.4	1205	9	US-10-084-994-2
9	395.4	11.6	2700	9	US-10-084-994-5
10	175	5.1	1686	9	US-10-084-994-6
11	165.8	4.8	1578	12	US-10-105-929-10
12	156	4.6	1464	9	US-10-152-661-255
13	156	4.6	1464	9	US-09-866-050A-255
14	156	4.6	1627	9	US-10-152-661-438
15	156	4.6	1627	9	US-09-866-050A-438
16	154	4.5	1633	9	US-10-152-661-73
17	154	4.5	1633	9	US-09-866-050A-73
18	154	4.5	2569	9	US-09-796-753-115
19	134.4	3.9	393	10	US-09-833-381-1076

20	129.6	3.8	1939	9	US-10-125-459-2	Sequence 2, Appl1
21	129.6	3.8	1939	9	US-10-067-761-4	Sequence 4, Appl1
22	129.6	3.8	1939	10	US-09-804-156-4	Sequence 2, Appl1
23	129.6	3.8	1939	10	US-09-946-633-2	Sequence 2, Appl1
24	89.2	2.6	3408	9	US-09-870-759-133	Sequence 139, App
25	89.2	2.6	3412	9	US-09-870-759-139	Sequence 139, App
26	85.8	2.5	8317	9	US-10-091-504-1279	Sequence 1279, Ap
27	85.8	2.5	8317	10	US-09-764-869-1279	Sequence 1279, Ap
28	82.8	2.4	2598	9	US-09-842-758-19	Sequence 19, Appl
29	82.8	2.4	2804	9	US-09-842-758-17	Sequence 17, Appl
30	64.2	1.9	540	9	US-10-073-865-44	Sequence 44, Appl
31	64.2	1.9	540	10	US-09-764-853-215	Sequence 215, App
32	64.2	1.9	540	10	US-09-764-898-61	Sequence 61, Appl
33	54	1.6	3914	12	US-10-044-090-148	Sequence 148, App
34	47.4	1.4	2005	10	US-09-895-072-18	Sequence 18, Appl
35	47.4	1.4	2005	10	US-09-986-552-18	Sequence 18, Appl
36	43.4	1.3	2739	9	US-10-091-504-1280	Sequence 1280, Ap
37	43.4	1.3	2739	10	US-09-764-869-1280	Sequence 1280, Ap
38	41.2	1.2	602	9	US-10-073-865-149	Sequence 149, App
39	41.2	1.2	602	9	US-10-073-865-150	Sequence 150, App
40	41.2	1.2	602	10	US-09-764-853-925	Sequence 925, App
41	41.2	1.2	602	10	US-09-764-853-926	Sequence 926, App
42	41.2	1.2	602	10	US-09-764-898-305	Sequence 305, App
43	41.2	1.2	602	10	US-09-764-898-306	Sequence 306, App
44	40.4	1.2	667	10	US-09-867-550-1911	Sequence 1911, Ap
45	40.2	1.2	756	9	US-10-123-155-392	Sequence 392, App

ALIGNMENTS

RESULT 1

US-10-092-390-1
Sequence 1, Application US/10092390
Publication No. US20030013865A1

GENERAL INFORMATION:

APPLICANT: Yu, Xuanchuan

APPLICANT: Miranda, Maricar

TITLE OF INVENTION: NO. US20030013865A1 Human BCF-Family Proteins and Polynucleo

FILE REFERENCE: LEX-0317-USA

CURRENT APPLICATION NUMBER: US/10/092,390

CURRENT FILING DATE: 2002-03-06

PRIOR APPLICATION NUMBER: US 60/275,013

PRIOR FILING DATE: 2001-03-12

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FASTSEQ for Windows Version 4.0

SEQ ID NO 1

LENGTH: 3423

TYPE: DNA

ORGANISM: homo sapiens

US-10-092-390-1

Query Match

Best Local Similarity 100.0%; Score 3423; DB 9; Length 3423;

Matches 3423; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCTATTCTTTGTAACATGCTGACCTTATTTGTTATGTTATGACCTGATG	60
DB	1	ATGCTATTCTTTGTAACATGCTGACCTTATTTGTTATGTTATGACCTGATG	60
QY	61	GGGACAGCATCACCCTGTAATCTTGAAGCCCTAATGTTAGCCACGGAAGCTAC	120
DB	61	GGGACAGCATCACCCTGTAATCTTGAAGCCCTAATGTTAGCCACGGAAGCTAC	120
QY	121	TCAGTACTGTCTCAGAGTATCATCCATCCCTTTGATCAATTTACTACAGAGCTGC	180
DB	121	TCAGTACTGTCTCAGAGTATCATCCATCCCTTTGATCAATTTACTACAGAGCTGC	180
QY	181	ACTGACATTTCTAACTGTTTAAATGACGCGGACAGAGTACGCTATGACAGCTAT	240
DB	181	ACTGACATTTCTAACTGTTTAAATGACGCGGACAGAGTACGCTATGACAGCTAT	240
QY	241	CGACATGGGAGAGACATATGATATAGCGGACAGTCTCATGTGCTGATTTATGAA	300

|||||
Db 241 CGACATGGGAGAGACATATGATAGCGCAAGCTCAGTGTCTCTCGATTTATGAA 300
Oy 301 AGGGGGAAATGTGTCCCCCACTGTGTATATGTATGTCATGTCGCTGATTTGCT 360
Db 301 AGGGGGAAATGTGTCCCCCACTGTGTATATGTATGTCATGTCGCTGATTTGCT 360
Oy 361 CCAAAACCTGTCAAGTGTAGCCCTGTGGGAGGAGCAACATGTCTCAGTGTGCGAT 420
Db 361 CCAAAACCTGTCAAGTGTAGCCCTGTGGGAGGAGCAACATGTCTCAGTGTGCGAT 420
Oy 421 GGTATACACTGTGGGTCCCCCACTGTACCAAGCGGTGTCCAGTGTCAAAAATGGGGCTGTGTC 480
Db 421 GGTATACACTGTGGGTCCCCCACTGTACCAAGCGGTGTCCAGTGTCAAAAATGGGGCTGTGTC 480
Oy 481 AACCCATCAACCGGGGCTTGGCACTGTGTCTGCGGGCTTCCGGGGCTGGCGTGGAGAG 540
Db 481 AACCCATCAACCGGGGCTTGGCACTGTGTCTGCGGGCTTCCGGGGCTGGCGTGGAGAG 540
Oy 541 CGCTGTAGCAGGAGCACCCTATGTATACGACTGTCTATCAGAGATGCCAGTGCAGATGGA 600
Db 541 CGCTGTAGCAGGAGCACCCTATGTATACGACTGTCTATCAGAGATGCCAGTGCAGATGGA 600
Oy 601 GCCACCTGCGACACAGTCAAGGGGAAATGCGCTGCGCACAGAGATACCGGAGCCTTC 660
Db 601 GCCACCTGCGACACAGTCAAGGGGAAATGCGCTGCGCACAGAGATACCGGAGCCTTC 660
Oy 661 TGTGAGGATCTTTTCTCCTGCTGTAACATGTGTCCACAGTGTGAGCAGAGATGCCCTTGT 720
Db 661 TGTGAGGATCTTTTCTCCTGCTGTAACATGTGTCCACAGTGTGAGCAGAGATGCCCTTGT 720
Oy 721 CAAAATGAGAGAGTGTGTCTATCAGTCACTGTGAGAGATGTCTTCCCTTGTGCTGATG 780
Db 721 CAAAATGAGAGAGTGTGTCTATCAGTCACTGTGAGAGATGTCTTCCCTTGTGCTGATG 780
Oy 781 GGCACAGTGTGTGTCAAGCCTTGCCTGGAGGGTGGCTTGGAAAGAACTGTGCCAAGAA 840
Db 781 GGCACAGTGTGTGTGTCAAGCCTTGCCTGGAGGGTGGCTTGGAAAGAACTGTGCCAAGAA 840
Oy 841 TGCCAGTGCCTAATGAGAGGAGCAGTGTATGCTGCCACAGGCAACATGTATGCACTGCA 900
Db 841 TGCCAGTGCCTAATGAGAGGAGCAGTGTATGCTGCCACAGGCAACATGTATGCACTGCA 900
Oy 901 GGAATACACAGGAGGAGAGCTGTCCAGAGATGATGTCTGTGGAGCTATGCGCTTCTGT 960
Db 901 GGAATACACAGGAGGAGAGAGCTGTCCAGAGATGATGTCTGTGGAGCTATGCGCTTCTGT 960
Oy 961 GCTGAGACCTGTCCAGTGTGTACAGGAGAGTGTACCAAGTGTGAGCGGCGCATGCTTC 1020
Db 961 GCTGAGACCTGTCCAGTGTGTGTACAGGAGAGTGTACCAAGTGTGAGCGGCGCATGCTTC 1020
Oy 1021 TGTGAGACAGGCTTGTGTGGGAGCGGTGCGAAGCAGCGCTGTCTGTGAGGGGCTGTAC 1080
Db 1021 TGTGAGACAGGCTTGTGTGGGAGCGGTGCGAAGCAGCGCTGTCTGTGAGGGGCTGTAC 1080
Oy 1081 GGCATCAAAATGTGACAAACGCTGTCTGTCCCTGCACTTGGAAAACATCATAGTGTACCCC 1140
Db 1081 GGCATCAAAATGTGACAAACGCTGTCTGTCCCTGCACTTGGAAAACATCATAGTGTACCCC 1140
Oy 1141 ATGTCTGAGAGTGTGTGCTGTGCAAGCGGGGTGTGAGAGCTCTAAGTAAATGAGACATGT 1200
Db 1141 ATGTCTGAGAGTGTGTGCTGTGCAAGCGGGGTGTGAGAGCTCTAAGTAAATGAGACATGT 1200
Oy 1201 TCTCTGATTTCTACGGGGAAAGCTTGGCAGAGATGTGCAAGCTGCCAAAATGGGGAGAG 1260
Db 1201 TCTCTGATTTCTACGGGGAAAGCTTGGCAGAGATGTGCAAGCTGCCAAAATGGGGAGAG 1260
Oy 1261 TGTGACAGTGTGACTGAGAAAGTGCAGCTGTGCCCCAGAGATTCAAAAGAAATGACTGCTCT 1320
Db 1261 TGTGACAGTGTGACTGAGAAAGTGCAGCTGTGCCCCAGAGATTCAAAAGAAATGACTGCTCT 1320
Oy 1321 ACCCATGGCCCTTGGGAACTATGGGATTAACGTGTCTCTGCTGTGGCTGTAAAAAT 1380
Db 1321 ACCCATGGCCCTTGGGAACTATGGGATTAACGTGTCTCTGCTGTGGCTGTAAAAAT 1380

Db 1381 ACCCATGGCCCTTGGGAACTATGGGATTAACGTGTCTCTGCTGTGGCTGTAAAAAT 1380
Oy 1381 GATCAGCTGTGTCTCTCTGTGAGAGCGGTCTTGTACTGTCAAGGAGCGGTGCAGCGGGT 1440
Db 1381 GATCAGCTGTGTCTCTCTGTGAGAGCGGTCTTGTACTGTCAAGGAGCGGTGCAGCGGGT 1440
Oy 1441 GACTGTCTCATAGAGTGTCCAGTGTGAGCATGTGGGGCTTGTGTAACTTAACATGCGAG 1500
Db 1441 GACTGTCTCATAGAGTGTCCAGTGTGAGCATGTGGGGCTTGTGTAACTTAACATGCGAG 1500
Oy 1501 TGTCTCAACGGGGAGCGCTGCACACACCTGTGACAGCGGACGTGCAGCTGTGAGATG 1560
Db 1501 TGTCTCAACGGGGAGCGCTGCACACACCTGTGACAGCGGACGTGCAGCTGTGAGATG 1560
Oy 1561 CGCGGGAGAAATGCGAACTTCCCTGTCCAGAGATGGACAGTGGGCTGAATGTGTGAG 1620
Db 1561 CGCGGGAGAAATGCGAACTTCCCTGTCCAGAGATGGACAGTGGGCTGAATGTGTGAG 1620
Oy 1621 CGCTGCGACTGTGACCCACAGCAGATGTGCTGCCACCTTACACAGGGCCATTGGCGCTCTC 1680
Db 1621 CGCTGCGACTGTGACCCACAGCAGATGTGCTGCCACCTTACACAGGGCCATTGGCGCTCTC 1680
Oy 1681 CCGGAGTGTGAGGTGTGCACTGTGACAGGAGTGTGTGCTGAGGAGAGCTGGGGCCCAAC 1740
Db 1681 CCGGAGTGTGAGGTGTGCACTGTGACAGGAGTGTGTGCTGAGGAGAGCTGGGGCCCAAC 1740
Oy 1741 TGTCTCCTGTGCTGTACTGTATTAATGGGGCTTATGCTTCCCTGTATGATGTCATGTC 1800
Db 1741 TGTCTCCTGTGCTGTACTGTATTAATGGGGCTTATGCTTCCCTGTATGATGTCATGTC 1800
Oy 1801 GAGTGTGACACAGGCTTCCAGAGCAGATGTGTAGAGATGTCTTCCCTGTGCTTTTAT 1860
Db 1801 GAGTGTGACACAGGCTTCCAGAGCAGATGTGTAGAGATGTCTTCCCTGTGCTTTTAT 1860
Oy 1861 GGGCATGCTGTGACAGCAGACATGTGCCACAGTGTGTACAGCAGCGGGCCCTGTGCACAC 1920
Db 1861 GGGCATGCTGTGACAGCAGACATGTGCCACAGTGTGTGTACAGCAGCGGGCCCTGTGCACAC 1920
Oy 1921 ATCACGGGCTGTGTGACTGTGCTTGTCTTACAGAGCGGCTTGTCAATGAATGTGT 1980
Db 1921 ATCACGGGCTGTGTGACTGTGCTTGTCTTACAGAGCGGCTTGTCAATGAATGTGT 1980
Oy 1981 CCCAGTGGCAGATTTGGGAAAACGTGTGAGAGATTTGTACCTGTGACCAACAGGAAC 2040
Db 1981 CCCAGTGGCAGATTTGGGAAAACGTGTGAGAGATTTGTGTACCTGTGACCAACAGGAAC 2040
Oy 2041 TGTAAACCCATTTGACATCTTGTGACTGTATACCCGGTGGATTTGGCAGTGTGCT 2100
Db 2041 TGTAAACCCATTTGACATCTTGTGACTGTATACCCGGTGGATTTGGCAGTGTGCT 2100
Oy 2101 CAACCATGTCCACTGTGCCACTGGGGCCCAAACTGTATCCACAGTGTCAACTGCCATAT 2160
Db 2101 CAACCATGTCCACTGTGCCACTGGGGCCCAAACTGTATCCACAGTGTCAACTGCCATAT 2160
Oy 2161 GGACCTTGTGACAGCGCTGTGAGATGGGAAATGTAAATGCACTCTGTGTGACAGGCTC 2220
Db 2161 GGACCTTGTGACAGCGCTGTGAGATGGGAAATGTAAATGCACTCTGTGTGACAGGCTC 2220
Oy 2221 TACTGTCACTAGAGATGTCTCTAGGGTTTATGAAAAGATTTGTGCACTGATATGCCAA 2280
Db 2221 TACTGTCACTAGAGATGTCTCTAGGGTTTATGAAAAGATTTGTGCACTGATATGCCAA 2280
Oy 2281 TGTCAAAAAGAGCTGTGAGTGTGAGACCAATTTCTGGGAGTGTACTTGGCCGACTGATTC 2340
Db 2281 TGTCAAAAAGAGCTGTGAGTGTGAGACCAATTTCTGGGAGTGTACTTGGCCGACTGATTC 2340
Oy 2341 ATGGGAGCGGACTGTGAGCAGAAATGCTTCAAGAAATATGCTATGTGCTGTGCGCAG 2400
Db 2341 ATGGGAGCGGACTGTGAGCAGAAATGCTTCAAGAAATATGCTATGTGCTGTGCGCAG 2400
Oy 2401 ATATGTATGTGTGCAACACTGTCCAGCGGACAGACATCACTGGAGCTGTACTGAGC 2460
Db 2401 ATATGTATGTGTGCAACACTGTCCAGCGGACAGACATCACTGGAGCTGTACTGAGC 2460


```

PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/259,388
PRIOR FILING DATE: 1999-02-26
PRIOR APPLICATION NUMBER: 60/122,458
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: 09/312,359
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/336,536
PRIOR FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: 09/342,687
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 09/345,464
PRIOR FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: 09/365,164
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/399,723
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 09/409,634
PRIOR FILING DATE: 1999-09-30
PRIOR APPLICATION NUMBER: 09/471,179
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/474,071
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/474,072
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/514,010
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: 09/516,745
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/572,002
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: 09/597,993
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 09/599,596
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/630,334
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 09/606,565
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/606,317
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/665,666
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: 09/677,751
PRIOR FILING DATE: 2000-09-30
NUMBER OF SEQ ID NOS: 162
SEQ ID NO 123
LENGTH: 3567
TYPE: DNA
ORGANISM: Rattus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (925)...(2832)
US-09-796-753-123

```

```

Query Match 17.8% Score 608.4; DB 9: Length 3567;
Best Local Similarity 56.7%; Pred. No. 5.8e-188;
Matches 126; Conservative 0; Mismatches 901; Indels 37; Gaps 4;

```

```

QY 67 GCATCACTCTGTAATCTTGAAGACCTTAATGTGTAGCCACTGGGAAGACTACTGAGT 126
DB 236 GCTGAACACTCAATCCATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 295
QY 127 ACCTGCAAGAGTCAATCCACATCCCTTT---GATCAATTTACTACAGAGAGTCACT 183
DB 296 ACACACTAAGAGTCCACCTTCCCTCCCTGAGCTGCCCCCAGCCAGTCTGCGACAG 355
QY 184 GACATTTAAACTGTTTAAATGACGCGCACAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 243
DB 356 CCTTGGGAAGACCCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 415
QY 244 CATGGGAGAGACTATGATAGCGCAAGTCTCACTGTTGCTGCTGCTGCTGCTGCTGCTGCT 303

```

```

DB 416 CAGTGTGTAAGATGAGTCCCGCCACGCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
QY 304 GGGGAATGTTGTTCCCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 363
DB 476 AGTGAAGCTGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 535
QY 364 AACACCTGTCAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 423
DB 536 AATCGGTGCCAGTGTCCACACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 595
QY 424 GATCACTGGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 483
DB 596 GGAATGGGGAACACAGTGTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 655
QY 484 CCGATACCGGGGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 543
DB 656 CCGAGAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715
QY 544 TGTGACAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 603
DB 716 TGCCCCGATGGCCACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772
QY 604 ACCTGGACACGTCACGGGGGGAATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663
DB 773 TCTGTGACCCCGGGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832
QY 664 GAGGATCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 723
DB 833 CTGATGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 870
QY 724 AATGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 783
DB 871 AATGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 930
QY 784 ACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 843
DB 931 GTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 990
QY 844 CAGTGCATTAATGAGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 903
DB 991 CTTGTCACAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1050
QY 904 TACACAGGGGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 963
DB 1051 TATATGGGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1110
QY 964 GAGACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1023
DB 1111 GAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1170
QY 1024 GAAGCAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1083
DB 1171 GAACATGGCTTCAACAGGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1230
QY 1084 ATCAATGTGACAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1143
DB 1231 CTGAGTGCACAGATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1290
QY 1144 TCTGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203
DB 1291 CACGGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1350
QY 1204 CCTGATTTAGGGGAGCTTCCGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1263
DB 1351 CAGGACACGACGAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1410
QY 1264 GACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1323
DB 1411 CTGCGGACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1470
QY 1324 CCAATGCTTCTGGAACCTATGGGATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1383
DB 1471 CTTTGTCCACCTTAACACTTATGGGATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1530

```


Db 182 CTCGCCAGTAGATGGCTCTGTAACCTGCAAGGAGGTGGCAGGGCTTGACCTC 241
QY 1452 CAGATGTCCAGTGGACATGGGGCTTTGGCTGTAACTTAATGCCAGTGGCTCAAGG 1511
Db 242 GCCATGTCCAGTGGAGCTGGGGCTTGAACATGCAACGAGACCTGACCTGTGCCAATGG 301
QY 1512 GGGAGCTGCACACACCTGGAGGAGACCTGCACGTGTGCATGTGATGGCGGGAGAGA 1571
Db 302 GGCACCTGCACACCCCATATGAGGGCTCTGCTCTGTGCACCTCTGGCTGGAGAGAC 361
QY 1572 ATGCCAACTTCCCTGCCAGGATGGACAGTACGGGCTGAACTGTGTGACCGCTGCAGCTG 1631
Db 362 CTGTAGAGCTGCTTGGCCGATGGCAATTTGGGCTGAACATGCAAGTGAACATGTGACCTG 421
QY 1632 CAGCCAGCAGATGGCTGGCACCCTACACGGGGCCATTGGCCCTCTCCCGGATGTGTC 1691
Db 422 CAGCCATGTGTATGTGATGTGACCCCGTACAGGCACTGCTCTGCTGGCGATGAGAC 481
QY 1692 AGGTGTCACTGTGACAGCGGTGTGTGAGGAGACCTGGGGCCCAACTGTCTCCCTGCC 1751
Db 482 AGGCATCCGCTGTGACAGAGAGCTGTCCACTGGCGCTGGGGCCCAACTGTCTGTCTTC 541
QY 1752 CTGTACTGTATAAAATGGGGCTTCACTCTCCCTATGATGTGCAATGTGCACTGTGCACC 1811
Db 542 CTGCAAGCTGTGAGAAATGAGGCTCTGCTCCCAAGAGGATGGGAGCTGCGAGTGTGCC 601
QY 1812 AGGCTTCGAGGACCACTGTGACAGAGATGCTGCTCCCTGTTTATAGGCAATGCTG 1871
Db 602 TGGCTTCGAGAGACCTTATGCAAGAGATGTGCCCCCTGGGTTCTATGGCAGCGCTG 661
QY 1872 CAGCCAGACATGCCACAGTGGCTTACAGACAGCGGGCCCTCCACACATCAGCGGCT 1931
Db 662 GCCCAGGACATGCCCTCTGCTGTGCAAGAGAGAGAGCGCTCCACACATCAGCGGCT 721
QY 1932 GTGTGACTGCTTGGCTGCTTACAGAGCGCCCTGTGCAATGAG 1975
Db 722 CTGTAGTGTCTCCAGAGATCTCTGTGAGCTCTGTCAACCAAG 765

RESULT 9

US-10-084-994-5
; Sequence 5, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO11PICI
; CURRENT APPLICATION NUMBER: US/10/084,994
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 5
; LENGTH: 2700
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-084-994-5

Query Match 11.6%; Score 395.4; DB 9; Length 2700;
Best Local Similarity 52.0%; Freq. No. 3.5e-118;
Matches 1019; Conservative 0; Mismatches 921; Indels 21; Gaps 5;

QY 407 CCAGTGGCTGCAGTGTATCATCTGGGGTCCCACTGACACGCGGTGCAGTGCATAAA 466
Db 15 CCCAGCGTCCGACAGAGCTGTATGTGCTCCAGCTGCCAGCAAGGTCTTTTGTGCA 74
QY 467 ATGGGGCTCTGTGCAACCCCATCACCGGGGCTTGGCACTGTCTCGGGCTTCCGGGCT 526
Db 111 1 1 111 111 1111 11 11111 1 11 1

Db 75 ATATGGGCACTGCCACCCAGCCACCGGACATGCACTGTGTCGCCCGGGTGGACCGCT 134
QY 527 GGGCTGCGAGGACCGCTGTGTAGCAGGGCACCCTATGTATAGCATGTATCAGAGATGCC 586
Db 135 TTAGCTGCGAGAGACCTGTGTATCTAGTGGCACTGGGGACCTTACAGAGCCACCTTCA 194
QY 587 AGTGCAGATAGGAGCC---ACCTGCAGACCACTACAGGGGAGAGCCGCTCCACAG 643
Db 195 ACTGCAGCGCTGTGGCCACGGAGCTGTGATGCTATACGGGCGCTGTGTCTGTGTAGGCTG 254
QY 644 GATACACGGAGACCTTCTGTGTAGGATCTTGTCTCTCTGTGTAAAGATGTCCACAGTGTG 703
Db 255 GCTATGTGGGGCCCGGGGTGGAGAGAGTGTGTCGCCAGAGGCCACTTGGGGCCGCTGTG 314
QY 704 AGCAGAGATGCCCTTGTCAAAATGAGAGATGTGTATCATCACTCACTGAGAAATGCTCTT 763
Db 315 AGCAGCGGTGCCAGTGTGACAGATGAGAGAGCTGTGTACACACTGACGGGGCTCTCAGCT 374
QY 764 GCGCTTGTGGCTGTGATGGGACACAGTGTGTGTGACCTTGTGCCGAGGGTGTCTTGGAA 823
Db 375 GCCCGGGCGCTGTGAGGGGACCTTGTGAGACATGTCTGCCCGCGCTTCTTGTGAT 434
QY 824 AGAAGTGTCCAGAGATGCCAGTGCATTAATGAGAGGAGCTGTATGCTGTGCACAGGCC 883
Db 435 TGGACTGTGCAAGTGTGCTGCAACTCACCGCGAGTGTGCTGTGATGCTGTGATGTGCT 494
QY 884 AATGTATTTGCAGTGTCCAGATATCACAGGGAAGAGGTGCCAGGATATAGTGTCTGTGGGA 943
Db 495 CTTGCTTGTGCGCCGCTGGGCGCCGCGGGGCTGTGTCGACAGACTGTCCAGGCCACA 554
QY 944 CTTATGGGCTTCTGTGTGTGAGCTGTGCAGTGTGTGTGTCAGAGAGGAATTTTACACAG 1003
Db 555 CTTAGGGGACAAATTTGCAAGCAGGCTGTGCTGCTTTTAAAGGGGCTCTGTGACCTG 614
QY 1004 TGAAGGCGCATGCTCTGTGTGAGCAGGCTTGTGTCGGAGCGCTGTGCAAGCAGCTGT 1063
Db 615 TCCAGGGGACAGTGCACATGTGCTCCCTGTGATGAGG---GCCCTGTGCTGTGAGGCT 671
QY 1064 GTGCTGAGGGGCTGTACGGCATCAATGTGCAAAAGGCTGCCGCACTGTGAAACA 1123
Db 672 GCGCTGCGGCTGTGTACGGGACCAACTGT-----CGGATTTCTGCTGTGCCAAGAG 725
QY 1124 CTATAGCTGTGACCCCATGTGTGTGAGAGTGTGCTGCAAGCGGGCTGTGTGAGACTCT 1183
Db 726 GAGGAGACCTGTGACCTGTGTGAGGCACTGTGCGTGTGCCAAGAGGCTGTGGGGCTGTG 785
QY 1184 ACTGTATGTAGACATGTTCTCTGTGATTCACGSGGAAGCTTGCACAGATGTGCACT 1243
Db 786 CTTGTGAGAAAGAGTGTCTCCCGGAGCTGTGAGCTGTGCGGACACAGCGCGGT 845
QY 1244 GCCAAATGGGCACTGTGTGACATGTGTGACGAAATGTGACCTGTGCCAGATTC 1303
Db 846 GCTCAAGCGGGGCTGTGTGTGACCGGACACAGCGGGCTGTGCTGTGCCAGCGGCTGTGA 905
QY 1304 AAGGAATGACTGTCTTACCCCATGCTTGTGGAACCTATGAGGATTAATGTTCTCTCTC 1363
Db 906 CTGGGGACAAGTGTGAGAGCCCTGTGCGGGGCTGTGTGTGGAAGGCTGTGTGCCAGC 965
QY 1364 GCTGTGCTGTAAAAATGATGTGCTGTCTCTGTGTGAGAGGCTGTGTACTGTCAAGG 1423
Db 966 GCTGTAGCTGTGCTGTGCGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1025
QY 1424 CAGGCTGTGACAGGGGTGTGACGTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1483
Db 1026 CTGCTTACATGTGCTGTGCGGCTGTGCGAGAGGCTGTGCCACCGGCTTGTGGGAGACT 1085
QY 1484 GTAACTTAACATGTGCAAGTGTCAACGGG---GAGAGCTGTCAACACCTGTGAGAGGACT 1540
Db 1086 GTGGCAGATGTGTGCAAGTGTGCTGTGAGAAACCGGCTGTGCCACCTGTGCCACCGGACT 1145
QY 1541 GCAGGTGTGACCTGTGATGTGCGGGGAGAAATGTGGAATCTTCTGTGCCAGATGTGACGT 1600
Db 1146 GCTCATGT 1205

```

QY 1601 ACGGCTGAAGTGTGTCAGCGCTGCGACATGCGACGCGAGATGCTGCCACCTACCA 1660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1206 ATGGGCCAGGCTGTGAACAGCTGTGGGTCTCAACGGGGGCTCCTGTGATCGGCGCA 1265
QY 1661 CGGGGCAATTGGCGCTGCCGCCGAGTGTAGAGTGTCCACTGTGACAGCGCTGTGCTG 1720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1266 CGGGGGCTGCGCTGCCGCCACTGTGGTCTCTGGGAGCGAGTGTCAACCTCACTGTCGCG 1325
QY 1721 AGGAGCCTGGGGCCCCCACTGCTCCCTGCGCTCTACTGTAAATAATGGGCTTCATGCT 1780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1326 AGGGCCCTGTGGGCCCAACTGCGACCGACGATGTGGGTGTGGGAGGGGGCGGCTGGCG 1385
QY 1781 CCCCTGATGATGCAATGTGGAGTGTGCACAGGCTTCGAGGACCACTTTGTCAAGGA 1840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1386 ACCCTGTGACCGGCACTGCTGCTGCGCCCGGGGAGAGCGCGCTGCTGAGCGAG 1445
QY 1841 TCTGCTCCCGGTTTTATGGGCAATGCGTGCAGCCAGACATGCCACAGTGGCTTACA 1900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1446 GCTGCCCCAGAACCGGTTTGGCTGGCGTGGCGAGCACACCTGCTCTCGAGAAAT--- 1501
QY 1901 GCAGCGGGCCCTGCGCACACATCACCGGCTGTGTGACTCTTGGCTTCACAGGCG 1960
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1502 -GGGGGCTGTGGCCAGCGGACGAGGAGGAGGAGCTGCTGTGGGCTGGGCTGAGCGGGCG 1559
QY 1961 CCCCTGCAATGAAGTGTGCTCCAGTGGCAGATTTGGGAAAACTGTGCAAGGAATTTGA 2020
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1560 GGCACCTCGAGCTGGCTGTGCTGCCCTGCGGCTACGAGGCGCGCTGCATCTGGAGTGTCT 1619
QY 2021 CTTGACCAACAGAGAACCTGTAAACCCATGTGACAGATCTGTGACATGTTACCCGGTT 2080
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1620 CTTGCCCAACAGAGAACCTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1679
QY 2081 GGATTTGAGTGAAGTGTCTCAACATGTCCACCTGCGGAGGAGGAGGAGGAGGAGGAGG 2140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1680 TCTATGAGCCAGGCTGTGGAGAGACCCCTGTGCCCTGTGCTTCAGAGGGGCTGGCGCAGG 1739
QY 2141 ACAGGTGCAACTGCGCATTAATGAGAGCTTGTGCGAGCGCTTACGATGGGAATGTAATGCA 2200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1740 GATTGTCTGTGTCACAAATGAGAGCCCTGCGAGCCCAATCAGTGGCGGAGTGTGCGCC 1799
QY 2201 CTTCTGCTGGAGAGGCTGTACTGACACTCAGAGATGCTCTTATGGGTTTATGGGAAAG 2260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1800 CTTGCCGCTTCCAGCGGCACTTCTGTAGAGGGGGTGTGAGCCAGTTTATGGAGAGG 1859
QY 2261 ATTGTGCACTGATATGCAATGTCAAAACGAGCTGACCTGACACCAATTTCTGGGAGT 2320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1860 GGTGCCACAGCGCTGTGACTGTGAGGGGGGGGACCCCTGTGACCCCTGTACCGGCTCT 1919
QY 2321 GTACTTGGCGCACTGATTTCAATGGGAGCGGACCTGTGAGCAG 2361
Db 1920 GCCTTTGCCACAGGGGCGCTCAGGAGCACCCTGTAACTGTG 1960

```

```

; LENGTH: 1686
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-084-994-6

Query Match      5.18; Score 175; DB 9; Length 1686;
Best Local Similarity 51.6%; Pred. No. 4.8e-46;
Matches 477; Conservative 0; Mismatches 440; Indels 8; Gaps 3;

QY 1165 CCGGGCTGTGAGAGACTCTACTGTAAATGAGACATGTTCTCTCGATTCATCGGGGAAAGCT 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 CTTGCTTCACTGTGCTCGGCTGTGAGACAGGAGATGTCCCGGGCGGTATGGGCCAGGC 74
QY 1225 TGGCAGCAGATCTGCACTGCTCCCAAAATGGGGCAGACTGGA-CAGTGTGATGGAAGTG 1283
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 TGTGAACAGTGTGTGGGTCTCTCAAGGGGGGCTCCTGTATGATGCGGCCACGGGGGCGCTG 134
QY 1284 CACTGTGCCCCAGAGATTCAAGGAATTAAGTCTGT-CTACCCCATGCGCTCTGTGGAACTT 1342
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 CCGCTGCCCACTGGGGTTCCTCGGGAGCGAGTGTCAACCTGACCTGTGCTCCGAGGGCGCT 194
QY 1343 ATGGGATAAACTGTTCCTTCGCTGTGGGCTGTAAATAATGATGAGTGTGCTCTGTGG 1402
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 TCGGCCCAACTGACACCCAGCTGTGTGGGTGTGGGAGGAGGGGCGCTGTGACCTGTGA 254
QY 1403 ACGGGCTTGTACTTGTGCAAGGAGGCTGCGACGGGGTGTGACTGTCCATGAGATGCCA 1462
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 CCGGACACTGCTCTGTGCCCCCGGGGAGAGCCGGCGTCCACTGTGAGCGAGGCTGCCGCC 314
QY 1463 GTGGCAGATGGGCTTGTGCTGTAACTTAACATGCGAGTGTCTCAAGGGGGAGCTGTGA 1522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 AGAACGGTTTGGCTGTGGGTGTGAGACACACTGTCTCTCAAGAAATGGGGGCTGTGCC 374
QY 1523 ACACCCCTGAGAGGAGGACTGTGACGTGTGACCTGTGATGAGCGCGGAGAAATGCGAACTTC 1582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 ACGCCAGCAACGGCAGAGTGTCTGTGTGGCTGTGGGCTGTGAGGGGGGAGCTGTGAGAGTGG 434
QY 1583 CTTGCCAGATGACAGTGTGAGGGCTGAACTGTCTGTGAGCGCTGCACTGACGAGCAGCAG 1642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 CTTGCTCCCTGTGGGCTGTGAGGAGCGCGCTGTGCATGTGAGTGTCTCTGCAACAAACA 494
QY 1643 ATGGCTGTCCACTTACACAGGGGCTATGCGGCTGCCCGGGATGTGACAGTGTCCACT 1702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 GCACTGTGTGAGCTGTGACAGGGGCACTGCGGCTGTGCGGCCCCGCTTCTATGACGAGGCT 554
QY 1703 GTACAGCGGTGTGCTGTGAGGAGCGCTGGGGCCCAACTGCTCCCTGTGCTGTACTGTA 1762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 GCGAGACACCCCTGTCCCTGTGCTTCCAGGGGCTGTGCTGCAAGGGGTTGTGCGGTGTC 614
QY 1763 AAAATGGGGCTTCATGCTCCCTGTGATGAGGCACTGTGCGAGTGTGACACAGGCTTCGAG 1822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 615 AACATGGAGCCCTGTGAGACCCCATCATGATGCGGATGCTCTGCTGCGGCTTCACAG 674
QY 1823 GCACCACTTGTGAGAGATGTGCTCCCTGTGTTTATGGGCAATGCTGTGAGCAGACAT 1882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 675 GCGCACTTCTGTGAGAGGGGTGTGAGCCAGGTTTATGAGAGGGCTGTGACACAGCGCT 734
QY 1883 GCCCAGAGTGTGCTTACAGAGCGGCGCTGTGACACACATCAGCGGCTGTGTGACTGCT 1942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 735 G-----TGACTGTGAGGGGGGGGACCCGTGTATACCCCTGTACACGGTCTGTGCTTTGCC 788
QY 1943 TGGCTGTGCTTACAGAGCGGCTGTGCAATGAATGTGTGCCAGTGTGAGGAAAA 2002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 789 CACACAGGGGCTGTGAGAGCACCCTGTAACTGTGATGCAAGAAAGGGGCGAGTTGGGCGCCA 848
QY 2003 ACTGTGAGGAATTTGTACTGTGACCAACAAGGAACCTGTAAACCATTTGACAGATCTT 2062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 849 GCTGACCCCTGTGACTGTGAGTGTGGGGGTGAGGCTGTGACCTGTACGTAGTGGCAGT 908
QY 2063 GTGAGTGTAAACCCGGTGGATTTGG 2087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 909 GTCACTGTGTGATGTGCTACATGGG 933

```

```

RESULT 10
US-10-084-994-6
; Sequence 6, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-Like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PT011P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 6

```



```

Oy 731 GAGTGTGTCATCAGCTGACAGTGAGAAATGAGCTTGGCTTCTGGCGAATGGGCACAGTGT 790
Db 241 GCCTGTGTCTACGTACCAATGAGCACTGCTCTCTGCCCCCTGGGCTGTGAGGGGCCACACT 300
Oy 791 GTGTGTACGCTTGGCCCCGAGGGTGCCTTTTGGAAAGAAGACTGTGTTCCCAAGATGCCAGTGC 850
Db 301 GTGAGACAGCGCTGCGTGTGAGGGGGCGTAATGTGTCTGCTGCGCTCCGTGGAGTGTCTGTCTC 360
Oy 851 ATAAATGAGGGAGCTGTGATGCTGACCAACAGGCCAATGTCTAATTTGACAGTCCAGGATACAG 910
Db 361 AGAACAATGAGCAGCTGTGAGCCCAACCTCGGCGCCTTGCTGTGTGTGGCCCTGACTTCATAG 420
Oy 911 GGGAGAGGTGCAGAGTAATGTGTCGTGTGGACACTAATGGCTTCTCTGTGTGAGACT 970
Db 421 GTCAAGCTTGTGAAGACACCTGCTGCCGCGCTTCATGATGCTGGTTGCCAGAGATTT 480
Oy 971 GCCAGTGTGTCAACGAGGAGGAAAGTGTAAACAGCTGAGCGGCGGCACATGCTGTGAAGAG 1030
Db 481 GCGAGTGTCAACAGGGGGCTCCCTGTGACCTGTCTACGTGTGAGTGGCGGTGCTCTGCTGCTG 540
Oy 1031 GCTTTGTGGGAGCGCTGTGGAAGACAGCGCTGTGTCCTGAGGGGGCTTACGGCATCAAT 1090
Db 541 GCTTCCGTGGCCAGTGTCTGCGAAGAGGGG---GTGCAAGCCAGGCTTTTGTGAGATGGCT 597
Oy 1091 GTGCAAAACGCTGTGCTCTGCCACTTGTGAAACACTATAGCTGTACACCCATGTGTGAG 1156
Db 598 GCGTGCAGCAGCTGTGACTGTGCTGCCCGCAGGGTGTGCTC-----CTGTGTCATCAACGCGGCC 651
Oy 1151 AGTGTGCTGTCAACGCGGGGTGTGTGCAAGACTGTACTGTAAATAGACATGTTCTCTGGAT 1210
Db 652 TCTGCTTTTGCCACACAGGGGCGCGAGAAACATGATGACCTAGATTGTGACAGAGAGGCC 711
Oy 1211 TCTACGAGGAGAGCTTTGCCACAGATGTGCAGCTGCCAAATGTGGGGCAGTGTGACAGT 1270
Db 712 GCTTTGGGGCGGGGCTGTGCGCTCGGTGGTGTGATTTGTGGGGGTGGGGGTGACTGTGACGCCA 771
Oy 1271 TGACTGGAAGTGCACCTGTGCCCCACAGATTTCAAAGCA 1308
Db 772 TCAGTGGGAGTGCACCTGTGTGAGACAGTACACGGGA 809

RESULT 15
US-09-866-050A-438
; Sequence 438, Application US/09866050A
; Publication NO. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene .
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishnand D.
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods for Their Use
; FILE REFERENCE: 11000.10114U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 438
; LENGTH: 1627
; TYPE: DNA
; ORGANISM: Mouse
US-09-866-050A-438

```

```

Query Match 4.6%: Score 156; DB 9; Length 1627;
      Best Local Similarity 51.2%: Pred. No. 8.1e-40;
Matches 419; Conservative 0; Mismatches 390; Indels 9; Gaps 2.

QY 491 CCGGGGCTTGCCACATGTGCTGCGGGCTTCGGGGCTCGCCGAGACCCCTGTAGCC 550
      1 CCGGAGCCCTGATACGCCCTGCTGGATTCCTTGGGAGCCAGCATGACCTTGGCTGTCCAC 60

```

OY	551	AGGGACCTTGGTAAGCATGTCTCAGAGATGCCAGGATGCCAATATGAGACACCTGCG	610
Db	61	AGGGTCCTTGGGCCACGCTGTGGCCACGGTGTATCATATGGCGGCAAGGGGGCGCATGTG	120
OY	611	ACCACTGTACAGGGGGAATGCCCTGGCCACCAAGSATAACCGGAGCTTCTGTGAGATC	670
Db	121	ACCAAGTGTGGGGGAATTTGATATCTGTCTCCGGGAAGACGGGABGCCATTGTAGCGCG	180
OY	671	TTTGTCCTCGTAAACATATGTGTCCACAGTGTAGCAGATGCCCCCTGTCTAAATGGAG	730
Db	181	GCTGTCCCCAGAGACCGGTTTGGCAAGGGCTGTGAACACAACTGTCTGTCAAGSATAAGG	240
OY	731	GAGTGTGCATACGTTCACAGSAGSAAATGCTTTGCCCTTCTGGCTGTAGTGGGACAGTGT	790
Db	241	GCTGTGTATGCTACTACCAATGTGAGCTGTCTCTGCCCTCGGGCTGTGAATGGGGCCACT	300
OY	791	GTGTGTACACCTTGTGCCCGGAGGTGTGGTTTGGAAAGAACTGTTCCTCCAGAAATGCCAGTGC	850
Db	301	GTGAGACAGGCTGCGCTGCGTGGGGCGGTATGTGTGTGCTGCGCTCTGGAGTGTCTGTTC	360
OY	851	ATAATGAGGAGACTGTGATGCTGCCACAGGCCAATGTATTGTACAGTCCAGATTCACAG	910
Db	361	AGAACATATGACAGTGTGAGACCCACCTCCGGCGCTTGCTCTGTGTGCCCTGTCTATG	420
OY	911	GGGAGCGTGCAGSATAGTGTCTGTTGGGACATATAGGCGTCTGTGTGTGAGACT	970
Db	421	GTCAAGCTTTGAGAGACACTGCCCTGGCCGGCTTCCATGTGATCTGTGGACAGATTT	480
OY	971	GCCAGTGTGTCAACGAGGGAATGTTTACACGTGACGCGGCATAGCTCTGTGTGAACAG	1030
Db	481	GCAGGTGTCAACAGGGGGGTCCCTGTGACCTGTGATGAGTGGCGGGTCTGTGCCCTGTG	540
OY	1031	GCTTGTGGGAGCGCTGTGGAAGCACAGCCTGTGTCTGATGAGGGGCTTACGGCATCAAT	1090
Db	541	GCTTCCGTGTGCAGTTCTGTGGAAGGGG---GTGCAGGCAAGGCTTTTGTGAGATGTGCT	597
OY	1091	GTGACAACGSGTGCCTGTGCACCTTGTGAAACACTATATGCTGTCAACCCATGTCGTGAG	1150
Db	598	GCTGTCAACGTGTACTGTGCCCAAGGGGTGTGC-----CTGTATCTCCATCAAGGGCC	651
OY	1151	AGTGTGCTGCACAGCCGGGCTGTGTACAGACTGTACTGTATATGACATGTTCTCCGTGAT	1210
Db	652	TCTGCTCTTGTGCCACGAGGGCGGGGAGGAACCATGTGACTAGATTGTGAGAAAGAGCC	711
OY	1211	TCATCGGGGGAAGCTTGTCCACAGATGTGCAGCTGCCAAATGGGCGAGACTGTGACATG	1270
Db	712	GCTTTGGGGCGGGCTGTGCTCTGCTGCTGTGATTTGTGGGGGTGGGCTGCTGTGCCCA	771
OY	1271	TGACTGGAAGTACACTGTGTGCCCAAGGAATTCAAAGGA	1308
Db	772	TCAGTGGGCACTGTGTGTGAGACAGCTTACACGGGA	809

Search completed: May 9, 2003, 11:53:28
Job time : 467 secs

THIS PAGE BLANK (USPTO)

Db 318 GCCTGCAACACCCCTGGACGGGACCTGCACGTGTGCACCTGGATGGCGGGGAGAAATGC 377

1576 GAACCTCCCTGCAGATGGACGATGAGGCGTGAACCTGCTGAGCGCTGGACGCTGACG 1635
1576 |||||||
378 GAACCTCCCTGCAGATGGACGATGAGGCGTGAACCTGCTGAGCGCTGGACGCTGACG 437
1636 CAGCAGATGGCTGGACGCTGACGATGAGGCGTGAACCTGCTGAGCGCTGGACGCTGACG 1695
438 CAGCAGATGGCTGGACGCTGACGATGAGGCGTGAACCTGCTGAGCGCTGGACGCTGACG 497
1696 GTCCAGCTGACAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1755
498 GTCCAGCTGACAGGCTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 557
1756 TACTGTAAAAATGGGCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1815
558 TACTGTAAAAATGGGCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 617
1816 TTCCGAGGACACACTTGTGACAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1875
618 TTCCGAGGACACACTTGTGACAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 677
1876 CAGACATGCCACAGATGGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1935
678 CAGACATGCCACAGATGGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 737
1936 GACTGCTTGCCTGGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1995
738 GACTGCTTGCCTGGCTGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 797
1996 GGGAAAACTGTGAGGAAATTTGTAACCTGACACAAAGGAACTGTGTAACCTGTAAC 2055
798 GGGAAAACTGTGAGGAAATTTGTAACCTGACACAAAGGAACTGTGTAACCTGTAAC 857
2056 AGATCTTGTAGTGTACCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 2115
858 AGATCTTGTAGTGTACCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 917
2116 GCCCACTGGGGCCCAACTGATGACACAGTGAACCTGCAATGAGCTTGTGACG 2175
918 GCCCACTGGGGCCCAACTGATGACACAGTGAACCTGCAATGAGCTTGTGACG 977
2176 GCGTCAATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 2235
978 GCGTCAATGAGGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 1037
2236 TGTCTCTAGGCTTGTGAGGAAATTTGTAACCTGACACAAAGGAACTGTGTAACCTG 2295
1038 TGTCTCTAGGCTTGTGAGGAAATTTGTAACCTGACACAAAGGAACTGTGTAACCTG 1097
2296 GACTGCGACACATTTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAG 2355
1098 GACTGCGACACATTTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAG 1157
2356 GAGCAGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAG 2415
1158 GAGCAGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAGTGTGCTGCGGAG 1217
2416 AACCACTGACCTGGACACATCTGGAACCTGTAACCTGTAACCTGTAACCTGTAACCTG 2475
1218 AACCACTGACCTGGACACATCTGGAACCTGTAACCTGTAACCTGTAACCTGTAACCTG 1277
2476 GCGAGATGTGATCAAGCTGTGATATCAATGTTGGAATCTGAACTGTAACCTGTAACCTG 2535
1278 GCGAGATGTGATCAAGCTGTGATATCAATGTTGGAATCTGAACTGTAACCTGTAACCTG 1337
2536 AGTACTGTCTCCCTGCTGATCTGACAGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2595
1338 AGTACTGTCTCCCTGCTGATCTGACAGATGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1397
2596 GTCCAGTGTGCT 2644
1398 GTCCAGTGTGCT 1446

RESULT 2
US-09-130-491-10
Sequence 10, Application US/09130491
Patent No. 6416974
GENERAL INFORMATION:
APPLICANT: Holzman, Douglas A.
APPLICANT: Goodheart, Andrew D.J.
TITLE OF INVENTION: TANGO-71, TANGO-73, TANGO-74, TANGO-76, AND TANGO-83
FILE REFERENCE: 09404/041001
CURRENT APPLICATION NUMBER: US/09/130,491
CURRENT FILING DATE: 1998-08-07
EARLIER APPLICATION NUMBER: US 60/058,108
EARLIER FILING DATE: 1997-09-05
EARLIER APPLICATION NUMBER: US 60/054,961
EARLIER FILING DATE: 1997-08-06
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 10
LENGTH: 1578
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(1578)
OTHER INFORMATION: n = A,T,C or G
US-09-130-491-10

Query Match 4.88; Score 165.8; DB 4; Length 1578;
Best Local Similarity 96.48; Pred. No. 9.6e-41;
Matches 212; Conservative 0; Mismatches 3; Indels 5; Gaps 4;
US-09-130-491-10

QY 3209 CCAACAGGAATG-TCTATGAGTGTGAACCT-ACAGTGAAGTGTGT--CCAAGAGTATTC 3264
DB 5 CCAACAGGAATGTTCTATGAAAGTGAACCTAAGCTGAGTGTGTGTTCCCAAGAGTATTC 64
QY 3265 AGCAATAATGGGCTGCTCTCCC-AGATCCATATGACCTCCCAAGAAAGTCAACATCC 3323
DB 65 AGCAATAATGGGCTGCTCTCCC-AGATCCATATGACCTCCCAAGAAAGTCAACATCC 124
QY 3324 TTGTCAATATGACCTGCTGCGGAGTGTGAGGAGTGTGATCCCTTGAAGAGGAGCAG 3383
DB 125 TTGTCAATATGACCTGCTGCGGAGTGTGAGGAGTGTGATCCCTTGAAGAGGAGCAG 184
QY 3384 TGGAGGTAGCAGCAGCAACAGCAGCAGCAGCAGCAGTGAATGA 3423
DB 185 TGGAGGTAGCAGCAGCAACAGCAGCAGCAGCAGCAGTGAATGA 224

RESULT 3
US-09-188-930-255
Sequence 255, Application US/09188930A
Patent No. 6150502
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Strachan, Lorna
APPLICANT: Sleeman, Matthew
APPLICANT: Onrust, Rene
APPLICANT: Murison, James Greg
TITLE OF INVENTION: Compositions Isolated From Skin Cells
FILE REFERENCE: 11000.1011c1
CURRENT APPLICATION NUMBER: US/09/188,930A
CURRENT FILING DATE: 1998-11-09
NUMBER OF SEQ ID NOS: 348
SOFTWARE: PastSeq for Windows Version 3.0
SEQ ID NO 255
LENGTH: 1464
TYPE: DNA
ORGANISM: Mouse
US-09-188-930-255

Query Match 4.68; Score 156; DB 3; Length 1464;


```

: ADDRESS: Immunex Corporation
: STREET: 51 University Street
: CITY: Seattle
: STATE: Washington
: COUNTRY: US
: ZIP: 98101
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/06093
: FILING DATE: 19930625
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/905,600
: FILING DATE: 26-JUN-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Seese, Kathryn A.
: REGISTRATION NUMBER: 32,172
: REFERENCE/DOCKET NUMBER: 2609
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 587-0430
: TELEFAX: (206) 233-0644
: TELEX: 756822
:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4138 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 149..3523
: PCT-US93-06093-1
:
: Query Match 1.6%; Score 55.2; DB 5; Length 4138;
: Best Local Similarity 60.8%; Pred. No. 2.8e-06;
: Matches 90; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
:
: Oy 1131 CTGTACCCCAATGTGTGAGAGATGTCCTGCACAGCCGGGCTGTGCAGACTTACTGTAA 1190
: Db 985 CTGTCTCCCTTACACCCCTATGGGTGTCTGTGCCACAGGCTGGAAGGGCTCTCAGATGCAA 1044
:
: Oy 1191 TGAGACATGTTCTCTCGATTCCTACAGGGGAAAGCTTGCCAGCAGATCTGCAGCTGCCAAA 1250
: Db 1045 TGAAGCATGCGCACCCCTGCTTTTACGGGGCAGATGTGAAGCTTAGTGAGCTGCGACGCAACA 1104
:
: Oy 1251 TGGGGACAGATGTGACAGCTGTGACGCA 1278
: Db 1105 TGGGAGATGTGTATCGCTTCCAAAGGA 1132
:
: RESULT 7
: US-08-220-240A-4
: Sequence 4, Application US/08220240A
: Patent No. 5935291
: GENERAL INFORMATION:
: APPLICANT: Alltalo, Karl
: APPLICANT: Matikainen, Marja-Terttu
: APPLICANT: Partanen, Juha
: APPLICANT: Makiela, Tomi
: APPLICANT: Korhonen, Jaana
: TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE RECEPTOR
: TITLE OF INVENTION: TYROSINE KINASE AND USES THEREOF
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Marshall, O'Toole, Garstein, Murray & Borun
: STREET: 233 South Wacker Drive/6500 Sears Tower
: CITY: Chicago
: STATE: Illinois

```

```

COUNTRY: United States of America
ZIP: 60606-6402

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/220,240A
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/F193/00006
FILING DATE: 08-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/817,800
FILING DATE: 09-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/167,453
FILING DATE: 15-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 29151/31958
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3845 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
FEATURE:
NAME/KEY: CDS
LOCATION: 37..3450
IS-08-220-240A-4

```

Query Match	1.6%;	Score 54;	DB 2;	Length 3845;
Best Local Similarity	51.4%;	Pred. No. 6.2e-06;		
Matches 179;	Conservative 0;	Mismatches 160;	Indels 9;	Gaps 2.

QY	1709	GCGTGTGTGCTGAGGAGCGCTGGGGCCCCCACTGCTCCCTCGCCCTGC--	-TACTGTAAAA	1765
Db	674	GGGGTGTGGGGGCTGGGGGCTGGGGGCCAGCGCTGACCAAGAGAGTGCAGGTTGCTTAC		733
QY	1766	ATGGGGCTTCAATGCTCCCTCATATGATGAGCATCTGCGAGTGTGCACACAGGCTTCCAGAGCA		1825
Db	734	ATGGAGGTGTCTGCGCACACACCATGACGGCGAGATGTGTATGCCCTTCGCTTCACTGCGCA		793
QY	1826	CCACTTGTCCAGAGGATCTGCTCCCTCGTGTATTTATGGGCATTCGTCGACGCCAGACATGCC		1885
Db	794	CCCCGTGTGAACAGGCCCTGTGCAGAGAGGGCCGTTTTGGGCAGAGCTCCAGGAGCAGTGGC		853
QY	1886	CACAGTGGCTTCACACGACGAGGGCCCTCCACACATCACCGGCTGTGTGACTGCTTGC		1945
Db	854	CAGGATATTCAGCGCTCGGGGGCCCTCACCTTCTGCTCTCCAGACCCCTATGGCTGCTCTT		913
QY	1946	-----CTGGCTTCCACAGGGCCCTCTGCAATGAAGTGTGCCAGTGGCAGATTTGGGA		1999
Db	914	GTGCACTGTGCTGGGAGAGGAAGCCAGTGTCCAAAGAGCTTGTCCCTCGGTCAATTTTGGGG		973
QY	2000	AAAAGTGTGCAGGAATTTGTACCTGCACCAACAAAGGAACCTGTAAAC		2047
Db	974	CTGATTCGCGACTCCACGTGCAGTGTTCAGAAATGTGGCATTGTGACC		1021

RESULT 8
US-07-934-393B-1
; Sequence 1, Application US/07934393B
; Patent No. 5466596
GENERAL INFORMATION:

```

1  APPLICANT: BREITMAN, MARTIN L.
2  APPLICANT: DUMONT, DANIEL
3  APPLICANT: GRADWOHL, GERARD G.
4  TITLE OF INVENTION: TISSUE SPECIFIC TRANSCRIPTIONAL
5  TITLE OF INVENTION: REGULATORY ELEMENT
6  NUMBER OF SEQUENCES: 5
7  CORRESPONDENCE ADDRESSES:
8  ADDRESSEE: BERESKIN & PARR
9  STREET: 40 King Street West
10 CITY: Toronto
11 STATE: Ontario
12 COUNTRY: Canada
13 ZIP: M5H 3Y2
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/07/934,393B
22 FILING DATE: 25-AUG-1992
23 CLASSIFICATION: 435
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Kurydyk, Linda M.
26 REGISTRATION NUMBER: 34,971
27 REFERENCE/DOCKET NUMBER: 3153-64
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (416) 354-7311
30 TELEFAX: (416) 361-1398
31 INFORMATION FOR SEQ ID NO: 1:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 4175 base pairs
34 TYPE: nucleic acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: cDNA
38 HYPOTHEICAL: NO
39 ANTI-SENSE: NO
40 FRAGMENT TYPE: N-terminal
41 ORIGINAL SOURCE:
42 ORGANISM: Mus pahari
43 STRAIN: CD-1
44 DEVELOPMENTAL STAGE: Embryo
45 TISSUE TYPE: Heart
46 IMMEDIATE SOURCE:
47 CLONE: tek
48 POSITION IN GENOME:
49 CHROMOSOME/SEGMENT: 4
50 MAP POSITION: Between the brown and pmv-23 loci
51 FEATURE:
52 NAME/KEY: CDS
53 LOCATION: 124..3477
54 OTHER INFORMATION: /function- "putative transmembrane
55 OTHER INFORMATION: receptor"
56 OTHER INFORMATION: /product- "tyrosine kinase"
57 OTHER INFORMATION: /gene- "tek"
58 OTHER INFORMATION: /standard- "tyrosine kinase receptor protein"
59
60 JS-07-934-393B-1

```

Query Match	1.6%	Score 54;	DB 1;	Length 4175;
Best Local Similarity	50.6%	Pred. No. 6.5e-06;		
Matches 166;	Conservative	0;	Mismatches 150;	Indels 12; Gaps
QY	716	CTGTGCAAAATGAGGAGGATGTGTCTATCAGCTCACTGAGAAATGCTTCTGCGCTTGGCT	775	
Db	800	CTTGCAGAAACATGTGAGTGTGTCATGAAAGATATCCGGGAAATGATTCATTCCTTCCTGGT	859	
QY	776	GGATGGGCAACAGTGTGTGTCAGGCTTCCCGGAGGGTGGCTTTGGAAGAAACACTGTCC	835	
Db	860	TTATGGGGAGAACATGTGTAGAAACTTGTGAGCCGCAACACTTGGTGGCAGGACCTGTAAAG	919	
QY	836	AAGATGCGCACTGGCATTAATGAGGAGGAGCTGTGATG-----CTGCCACAGGCC	883	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 10:55:56 ; Search time 18.4936 Seconds

(Without alignments)
2556.719 Million cell updates/sec

Title: US-10-092-390-2
Sequence: 1 MVSLNSCLSFICLLCHMT.....SSPKQDSGSSSSSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1037	15.4	2524	1	NOTC_XENLA
2	1034.5	15.3	2536	1	NTC1_HUMAN
3	1028	15.2	2531	1	NTC1_MOUSE
4	1024	15.2	2531	1	NTC1_RAT
5	1014.5	15.0	2471	1	NTC2_HUMAN
6	998	14.8	2471	1	NTC2_MOUSE
7	993	14.7	2470	1	NTC2_MOUSE
8	987	14.6	2437	1	NTC1_BRARE
9	978.5	14.5	2703	1	NOTC_DROME
10	977.5	14.5	2318	1	NTC3_MOUSE
11	974	14.4	2321	1	NTC3_HUMAN
12	969.5	14.2	2319	1	NTC3_RAT
13	959.5	14.2	2003	1	NTC4_HUMAN
14	954.5	14.2	1064	1	FBP1_STRPU
15	951.5	14.1	1964	1	NTC4_MOUSE
16	916.5	13.6	4289	1	TENX_HUMAN
17	870.5	12.9	830	1	SREC_HUMAN
18	832.5	12.3	1213	1	JAG3_BRARE
19	789	11.7	1238	1	JAG2_HUMAN
20	775.5	11.5	2139	1	CRB_DROME
21	775	11.5	1242	1	JAG1_BRARE
22	769	11.4	1247	1	TENA_HUMAN
23	768	11.4	1247	1	JAG2_MOUSE
24	767.5	11.4	1246	1	TENA_PIG
25	757	11.2	1218	1	JAG1_HUMAN
26	745	11.0	1202	1	JAG2_RAT
27	744	11.0	1218	1	JAG1_MOUSE
28	739	11.0	1218	1	JAG1_MOUSE
29	736.5	10.9	3695	1	TENA_HUMAN
30	720.5	10.7	1808	1	TENA_CHICK
31	717.5	10.6	1801	1	LMB2_RAT
32	716	10.6	3106	1	LMA2_MOUSE
33	706.5	10.5	1790	1	LMB1_DROME

ALIGNMENTS

RESULT 1	ID	NOTC_XENLA	STANDARD	PRT	2524 AA.
AC	P21783				
DT	01-MAY-1991 (Rel. 18, Created)				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Neurogenic locus notch protein homolog precursor (XORCH protein).				
GN	XORCH				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_Taxid-8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE-90385285; PubMed-2402639;				
RA	Ciftman C., Harris W., Kintner C.;				
RT	"Notch, the Xenopus homolog of Drosophila notch.";				
RL	Science 249:1438-1441(1990).				
RN	[2]				
RP	REVISIONS TO 1759-1782.				
RA	Kintner C.;				
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.				
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.				
CC	-1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.				
CC	-1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.				
CC	-1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 6 ANK REPEATS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: M33874; AB02039.1; -				P19137 mus musculus
DR	PIR: A35844; A35844.				O61292 mus musculus
DR	HSP: P00740; IEDM.				P25391 homo sapien
DR	InterPro: IPR002110; ANK.				P18168 drosophila
DR	InterPro: IPR000152; Asx_hydroxyl.				P24043 homo sapien
DR	InterPro: IPR000561; EGF-like.				P55268 homo sapien
DR	InterPro: IPR000742; EGF-2.				O61001 mus musculus
DR	InterPro: IPR001881; EGF_CA.				O21313 caenorhabdl
DR	InterPro: IPR001438; EGF-IT.				P07942 homo sapien
DR	InterPro: IPR000800; Notch.				P02469 mus musculus
DR	PIR: PFO0023; ank. 6.				P14585 caenorhabdl
DR	PIR: PFO0008; EGF. 36.				O00174 drosophila
DR	PIR: PFO0066; notch. 3.				
DR	PRINTS: PR00010; EGFLOOD.				
DR	PRINTS: PR01452; NOTCH.				
DR	SMART: SM00248; ANK. 5.				
DR	SMART: SM00179; EGF_CA. 23.				

Query Match 15.3%; Score 1034.5; DB 1; Length 2556;
 Best Local Similarity 25.8%; Pred. No. 3.6e-53;
 Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

```

FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.

QY 94 CCRPFYEGEMCVPHCADKCYHGR-----IAPNTQCEBPGMGTCSSACDGDH 143
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 CALFE--SGPLCLTPDLNACLTPNCRNGCTDLTLLEYKCRCPGWSGKSCQQA----- 141

QY 144 WGPHTSRQCKNGALCPRTIGA--CHCAAGFRGRCE---DRCEOG---TYGNDCHQ- 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 --DPCASN--PCANGGCLLPFASTYICHCPSFHFPTGCQDYNEGQAKRLCRHGTCINE 198

QY 194 ---RC-----OQNGATC--DHYTGECRCPGYTGAFCE--- 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 VGSRYCVRATHGTGPNCRERYVPCSPPCQNGTCRPTGDTYHACALPFTGQNCENI 258

QY 223 DLCPRG--KHGPQC-----EORCP-----CQNGVCCHHTG- 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 DDCBGNCKNGACVGDVNTYNCPCPREMTQCTEDYDECQLMPCQNGTCCHNTHG 318

QY 252 --ECSPSGMGTGQ-----PCBGRGRKNC--SQEC-- 281
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 YNCVCVNMWTEDECSSENIDOCASACFHGATCHDRVASFYECBGRGTGLLCHLNDACIS 378

QY 282 --QCHNGGTCA--ATGQ--CHCSPGYTERCO--DECPVGYGVLAETCQCVNGKCY 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 379 NPNMGNSMCDTPNPNKALICTCPSPSYTPACSDVDECSLAN-----PCBLAGICI 430

QY 334 HVSQA--CLCAGRAGECEARLCEBGLYIKDKRC---PCHLENTHSCHPMGE--CA 386
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 431 NTLSFEQCLQGYTGPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477

QY 387 CKPGMSGLYC-----NE--TCSPGFYGEACO--QICS--C 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 CMPEYEGHCEVNTDECASSPCLNGLRCLDINEFQCPCPTGFTGHLCQDYDECASTPC 537

QY 416 QNGADC-----DSV--TGKCTCAPGRKG 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 KNGAKCLDGPMTYTCVCTEGYTGTHCEVDIDEDCPDCHYGSCKDGVATFTCLCPGYTG 597

QY 437 IDGCT-----PCPL-----GTVINS-----SRGCANDAVCS 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 HHCEININESSQPCRLNGTCQDPDMAYLCFLGTTGPNCEINIDPCASSPCDSGTCLD 657

QY 466 PVDG--SCTCKAGMHGVDGCSIR-----CPSGTWGFQCNL-----TC----- 499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 658 KIDYECAECERGYGYSKMSNIDECAGNCPCHNGTCEGJINGFTCRCEPETHDPTCLSEV 717.

QY 500 -----OCLNGACNWLTDG--TCTCAPGMRGKCEL----- 527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 718 NECSNPNFVHAGACRDLNGYKCDGPGMSGTNCIDINNECESNPNVNGTCKDMTSGIYC 777

QY 528 PCQCGTYLNCARCD-----CSHADOC--HPTTGH--CRCLPGMSGVHCDV---CABG- 575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 TCRREGFSQPCQTNINECASNPCLNKGTCTIDVAGYKNCNCLLPYTGACCEVVAAPCASP 837

QY 576 --RMGPNC-----SLPCYC-----KNGACSPDQICGCAFGPRTGTQRT-----CS 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 838 CRNGECQSQEDYESFSCVCTPAGAKGTCCEVDINECVLSPCRHGACQNTNHCXTRCQ 897

QY 617 PGFYGHRSQTCPOCVHSSGPCHH-----ITGLDCLPFTGALCNE----- 658
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 AGYSGRNCETDIDC--RPNPCNGNSCTDGINFAFCCLLPGRFOTFEEDINECASDPC 955

QY 659 -----VCSGRFGKNCAG--ICT--CTNNGTCMPIDR-----SCQCTPG 693.
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 956 RNGANCTDCVDSYCTCTCPAGFSGJHCENNTDCTESSCNGTCT--VDGINSFTCLCPG 1013
  
```

```

QY 694 WIGSDC-----SOP-----CPAHMGPNC---IHTCN---CHNGA 722
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1014 FTGSCQHYVNECBSRFLLEGTCODGRLHRCICPGTYTPNQNVLVHWCDSPPCKNGG 1073

QY 723 FC-----SAYDECKCTPGMTGLYCTQ----- 744
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1074 KCMQTHRYQ--RCBPCSGMTGLYCDVPSVCEVAARQGVAVARLQHGGLCVDAQNTHH 1131

QY 745 --RCELFGYGRKCALI---CQ-----CQNGADC--DHISG--QCTCTGFMGRHCQK----- 788
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 CRCQAGTYSYCEDLVDECSPPCQNGATCTDYLGYSCKCVAGYHGVNCSSEEDCLSH 1191

QY 789 -----CPSGTGYGCRQICD-----CLANSTC--DHITG 815
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1192 PCQNGTCLDLPNTYKSCSPGTGVCINEIVDNCNPPVDPVYSNPKCFNNGTCVDQYGG 1251

QY 816 --TCYCSPGMGKARGCDQAGVIIVGNLN 840
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1252 YSCTCPGFEVERCE-----GDVN 1270

RESULT 3
NTCL_MOUSE STANDARD: PRT: 2531 AA.
AC 001705: Q9QW58; Q99JC2; Q60007; Q61905; Q9R0X7;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1) (Motch A)
GN NOTCH1 OR MOTCH.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid:10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Embryo;
RX MEDLINE-93194170; PubMed-8449489;
RA Franco del Amo F., Gendron-Maguire M., Swatek P.J., Jenkins N.A.,
RA Copeland N.G., Gridley T.;
RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
RT homolog of Drosophila Notch.";
RL Genomics 15:259-264(1993).
RN [2]
RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
RC STRAIN=CD-1; TISSUE=Embryo;
RX MEDLINE-93050801; PubMed-1426644;
RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;
RT "Expression analysis of a Notch homologue in the mouse embryo.";
RN [3]
RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
RC TISSUE=Embryo;
RX MEDLINE-93048835; PubMed-1425352;
RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
RA Greenspan R.J., McMahon A.P., Gridley T.;
RT "Expression pattern of Motch, a mouse homolog of Drosophila Notch,
RT suggests an important role in early postimplantation mouse
RT development.";
RN [4]
RP SEQUENCE OF 1161-1547 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE-93178563; PubMed-8440332;
RA Lardelli M., Lendahl U.;
RT "Motch A and Motch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues.";
RN [5]
RP Exp. Cell Res. 204:364-372(1993).
RX SEQUENCE OF 1659-1673 FROM N.A.
RX MEDLINE-99364499; PubMed-10437788;
RA Lee J.S., Ishimoto A., Yanagawa S.I.;
  
```

RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3."
RL FEBS Lett. 455:276-280(1999).

RM [6]
RN SEQUENCE OF 1950-2201 FROM N.A.
RP MEDLINE-98029496; PubMed-9384671;
RX Messierle M., Folio M., Nehls M., Eggert H., Boehm T.?
RY "Dynamic changes in gene expression during in vitro differentiation of
RT mouse embryonic stem cells";
RL Cytokines Cell. Mol. Ther. 1:139-143(1995).

RM [7]
RN SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
RP MTOGENESIS OF 1651-ARG-GLN-ARG-ARG-1654.
RX MEDLINE-98318619; PubMed-9653118;
RY Logeat F., Bessia C., Brou C., Lebail O., Varrault S., Seldah N.G.,
RA Israel A.?
RT "The Notch1 receptor is cleaved constitutively by a furin-like
RT convertase";
RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).

RM [8]
RN PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
RP MEDLINE-21523956; PubMed-11518718;
RX Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.?
RY "Murine notch homologs (NL-4) undergo presenilin-dependent
RT proteolysis";
RL J. Biol. Chem. 276:40268-40273(2001).

RM [9]
RN POST-TRANSLATIONAL PROCESSING.
RP MEDLINE-21374376; PubMed-11459941;
RX Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Hongo T.?
RY "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).

CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (by similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation. May be involved in mesoderm
CC development, somite formation and neurogenesis. Involved in the
CC maturation of both CD4+ and CD8+ cells in the thymus.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(Tm) and a N-
CC terminal fragment N(Ec) which are probably linked by disulfide
CC bonds.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC intracellular processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
CC produced by alternative splicing.

CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
CC thymus. Expressed at lower levels in the spleen, bone-marrow,
CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
CC muscle, kidney and heart.

CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c.
CC by 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme
CC and endothelial cells, while much lower levels are seen in the
CC neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels
CC in the neuroepithelium. At 13.5 d.p.c. expressed in the surface
CC ectoderm, eye and developing whisker follicles.

CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(Tm) and a N-terminal fragment N(Ec). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane.
CC -1- PTM: Phosphorylated.

CC	-1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC	-1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC	-1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC	-1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
CC	-----
DR	EMBL; Z11886; CAA77941.1; -.
DR	EMBL; L02613; AAK14898.1; -.
DR	EMBL; K68878; CAAA8339.1; -.
DR	EMBL; AJ238029; CAB840733.1; -.
DR	EMBL; X82562; CAAS7909.1; -.
DR	HSSP; P00740; IEDM.
DR	MGD; MGI:97363; Notch1.
DR	InterPro; IPRO02110; ANK.
DR	InterPro; IPRO00152; Asx_hydroxy1.
DR	InterPro; IPRO00561; EGF-like.
DR	InterPro; IPRO00742; EGF_2.
DR	InterPro; IPRO01881; EGF_Ca.
DR	InterPro; IPRO01438; EGF_I1.
DR	Pfam; PF00008; EGF; 35.
DR	Pfam; PF00023; ank; 7.
DR	Pfam; PF00066; notch; 3.
DR	PRINTS; PRO0010; EGFBLOOD.
DR	PRINTS; PRO1452; NOTCH.
DR	SMART; SM00248; ANK; 3.
DR	SMART; SM00179; EGF_CA; 23.
DR	SMART; SM00001; EGF_Like; 11.
DR	SMART; SM00004; NL; 2.
DR	PROSITE; PS50088; ANK_REPEAT; 2.
DR	PROSITE; PS50297; ANK_REP_REGION; 1.
DR	PROSITE; PS00010; ASX_HYDROXYL; 22.
DR	PROSITE; PS00022; EGF_1; 34.
DR	PROSITE; PS01186; EGF_2; 27.
DR	PROSITE; PS01187; EGF_CA; 21.
KW	Receptor; Transcription regulation; Activator; Differentiation;
KW	Developmental protein; Repeat; ANK repeat; Egf-like domain;
KW	Transmembrane; Glycoprotein; Signal; Phosphorylation;
KW	Alternative splicing.
FT	SIGNAL 1 .. 18 POTENTIAL.
FT	CHAIN 19 .. 2531 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
FT	CHAIN 1711 .. 2531 NORTH EXTRACELLULAR TRUNCATION.
FT	CHAIN 1744 .. 2531 NORTH INTRACELLULAR DOMAIN.
FT	DOMAIN 19 .. 1725 EXTRACELLULAR (POTENTIAL).
FT	TRANSHEM 1726 .. 1746 CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 1747 .. 2531 EGF-LIKE 1.
FT	DOMAIN 20 .. 58 EGF-LIKE 2.
FT	DOMAIN 59 .. 99 EGF-LIKE 3.
FT	DOMAIN 102 .. 139 EGF-LIKE 4.
FT	DOMAIN 140 .. 176 EGF-LIKE 5.
FT	DOMAIN 178 .. 216 EGF-LIKE 6.
FT	DOMAIN 218 .. 235 EGF-LIKE 7.
FT	DOMAIN 257 .. 293 EGF-LIKE 8.
FT	DOMAIN 295 .. 333 CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 335 .. 371 CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 372 .. 410 EGF-LIKE 10.
FT	DOMAIN 412 .. 450 EGF-LIKE 11.
FT	DOMAIN 452 .. 488 EGF-LIKE 12.
FT	DOMAIN 490 .. 526 EGF-LIKE 13.
FT	DOMAIN 528 .. 564 CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 566 .. 601 EGF-LIKE 15.
FT	DOMAIN 603 .. 639 CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 641 .. 676 EGF-LIKE 17.
FT	DOMAIN 678 .. 714 CALCIUM-BINDING (POTENTIAL).
FT	DOMAIN 716 .. 751 EGF-LIKE 19.
FT	DOMAIN 753 .. 789 CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 791 827 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 829 867 EGF-LIKE 22.
 FT DOMAIN 869 905 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 907 943 EGF-LIKE 24.

Query Match 15.2%; Score 1028; DB 1; Length 2531;
 Best Local Similarity 25.7%; Pred. No. 8.7e-53;
 Matches 314; Conservative 83; Mismatches 286; Indels 538; Gaps 73;

OY 86 TMYRRKSOCCPFYESENGCV--PHCADKCVH-GRCI--APNTCCCPFGMGATNCSS- 137
 DB 121 TLTEKRCRSPGM--SGKSCQADPCASNPANGGCLPEFESSYICRCPGRHGPTRQD 178
 OY 138 -----ACDGDHMGPRHC-----TSRCCKNALCNP--- 162
 DB 179 VNECSQNPGLCRHGHCHNEISYRACCATHTGPHCELPPVCPSPSPONATCRPTD 238
 OY 163 ITGACHCAAGFRGWRCEDECEOGTYGNDCHORCOCONGATC-DHY-TGECRCPGTYGAF 220
 DB 239 TTHECACLPFGAGNCEENVYD-----DCPGR-NCKNGACVGYNTYTCRCRPEYTGQY 291
 OY 221 C-EDLCPPKRGPOCEQRCP--QONGVCHVYTG--ECSCPGMMGTVCQO----- 266
 DB 292 CTEDEV-----DEC-QLMPNACQNAAGTCHNTHGVCYCVANGMTGEDCSENIDOCASAA 343
 OY 267 -----PCPEGKFGKNC--SQEC--QCHNGTCA--ATGQ--CHCSFG 301
 DB 344 CROGATCHDRVAFYCECHGRTGLLCHLKHACISNPCEGSCNCTPNVNGRITCPCPG 403
 OY 302 YNGERQO--DECPVGTGVLCAETCCQVNGKCYHVSQA--CLCEAGPAGERCEARLCP 356
 DB 404 YNGPACSDVDECDLGAN-----RCRHAKCLNTLGSFECCLQGTGCGCELDV--- 453
 OY 357 BELYGIKCDKR--PCHLENTSHCPMSGE--CAKCPGMSGLYCNET----- 399
 DB 454 -----NECISNPC--QNDATCIDQIGEFQICIMPEGEVYCEINTDECASSPCLHN 502
 OY 400 -----CSPGFGEACO--QICS--CONGADC--DSVTGCTCAPKRGKD 438
 DB 503 GHCMDKIEFOCCCPKRGFNGHLQYDVBECASCPCKNKGKCLDGPNTYTCVTEGTYGH 562
 OY 439 CST-----PCPLGT-----YGINCSSRGC-----CKNDVACSPVD 468
 DB 563 CEVDIDECDDPCHYGSCKMDGVAFTCLQPGYTHHCTNTNNECHSQCRHGRGTQDD 622
 OY 469 GS--CTCKAGHGVDCSIR-----CPSGTWFGCNILTCQCLNGGACNTLDG--TCICA 517
 DB 623 NSYLCICLKGTGPNCEIMIDCASNPDCSGT-----CL-----DKIDYEACAE 667
 OY 518 PGMREKCEL-----PCQOSTYGLNCAERDCSHADGCHFT----- 554
 DB 668 PGTGSMCVNINDECAGSPCHNGTCEBDIAGFTC--RC--PRGYHPTCLSEVNECN 721
 OY 555 -----GHCR-----CLPGWSGVHD-----SYCAE 574
 DB 722 SNPCJHAGCRDLNGYKKCCAPAGMSGTNDINNCESNPCVNGGCKMTSGYCTGAE 781
 OY 575 GWRGPNC-----SLPCT-----CKNG 590
 DB 782 GSGSPCOTNINECASNPCLNOGTICIDVAGYKCNCPLEPYTGATGEVVLAPCATSPCKNS 841
 OY 591 ASCSPDDGI-----CECAPFRGTQO-----RISGGEFG 621
 DB 842 GYCKSEDEYESTCYCPTGMOGQCEVDINCEYKSPCRHAGSCQNTNGSYRCLCAQAGTYG 901
 OY 622 HRCQOTPCQVHSSGPCNH-----ITGLCDCLPGFTGALCNE----- 658
 DB 902 RNCESDIDCC--RPNCHNGSGCTDGINAFCDCLPGFGARCEBDEINECANPQONGAN 959
 OY 659 -----VPSGRFGKNCAG--ICT--CTNNGTCNPIDR--SCQCPKIGSD 698
 DB 960 CTDCVDSYTCCTCPVGFNGIHCENNTPDCYESSCFFNGGTC--VDGINSFCLCPGPGTGSY 1017

OY 699 C-----SQP-----CPRAHMGPNC--IHNCN---CHNGAFC--- 724
 DB 1018 COYDVEDDSRPLCHGTCCQDSYTYKCTCPGGYGLNCONLVWCDSPCKNGRCMQT 1077
 OY 725 -SAYDECKCPWTGLYC--TORCPLEFY--GRDCALICQ----- 760
 DB 1078 NNGY--HCECRSGWGVNCDVLSYSCFEVAARQKIDVTLCOHGGLCYDEBKHCHCQA 1135
 OY 761 -----QONGADC-DHISG--QCTCTGFMGRHCEQK----- 788
 DB 1136 GYTGSGCEDEVEDCSPNPCQNAATCTDYLGFSKCVAGYHGSNCESEINELCSQPCQNG 1195
 OY 789 -----CPSGTYGCGRQICD-----CLNNSNC-DHITG-TEYC 819
 DB 1196 GTCIDLTNSYKSCSPRGTVGVCIEINVDCHPPLDPASRSPKCFNNGTCVQDVGGYGTCTC 1255
 OY 820 SPFMKGARCDQAGYIVGNLNL 840
 DB 1256 PPGYGERCE-----GDVN 1269

RESULT 4

NTCL_RAT STANDARD; PRF: 2531 AA.
 AC 007008:
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
 GN NOTCH1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.A.
 RC TISSUE=Schwann cell;
 RX MEDLINE=9211383; PubMed=1764995;
 RA Weimaster G., Roberts V.J., Lemke G.;
 RT "A homolog of Drosophila Notch expressed during mammalian
 RT development.";
 RL Development 113:199-205(1991).
 RN [2]
 RP REVISIONS TO 1652-1653.
 RA Weimaster G.;
 RL Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP FUNCTION.
 RX MEDLINE=21094508; PubMed=11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate.";
 RL Neuron 29:45-55(2001).
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=93202015; PubMed=1295745;
 RA Weimaster G., Roberts V.J., Lemke G.;
 RT "Notch2: a second mammalian Notch gene.";
 RL Development 116:931-941(1992).
 RN [5]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21331789; PubMed=11438922;
 RA Irvan D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-Delta signaling system during brain
 RT development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.

FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 227 243 BY SIMILARITY.
 FT DISULFID 245 254 BY SIMILARITY.
 FT DISULFID 261 272 BY SIMILARITY.

Query Match 15.2%; Score 1024; DB 1; Length 2531;
 Best Local Similarity 25.8%; Pred. No. 1.5e-52;
 Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

QY 86 TMTRRSSQCCPFGYSEGENCV---PHCADKCVH-GRCIT---APNTQCCEPFGMGTCSS- 137
 DB 121 TLTLEYKRCRPPGW--SGKSCQOQADPCASNPCCANGGOCLEPFESSYICGCPFGFHPYCRD 178
 QY 138 -----ACDDHMGPHC-----TSRCKKNGALCNP----- 162
 DB 179 VNECSNPGLCRHGTCHNEIGSYRCACRATHRPHCELPYPFCSPPCONGTCTPRTGD 238
 QY 163 ITGACHCAAGFRGWRCEDEQGTGNDCHORCOQNGATC-DHV-TGBCRCPPTGTAF 220
 DB 239 TTHECACLPFGAQNCEENVN-----DCPGR-NCKNGACVDGVNTYMCRCPPETGQY 291
 QY 221 C-EDLCPRKHGRQCRP--CQNGVCHHTG--ECSCPSGMMGTVCQ----- 266
 DB 292 CREDV-----DEC-QLMPNACQNGTCHNSHGGYNCVGVNWTGEDSDNIDDCASAA 343
 QY 267 -----PCEGRFRGKNC--SOEC--QCHNGGTCD-ATGQ--CHCSPG 301
 DB 344 CROGATCHNRVASFYCECHGRGLLCHNDACISNPEGSNDTPNPKRITCICPFG 403
 QY 302 YGERCO--DECPVGTGVLAEITGQCVNGKCYHVSAG--CLCBAGFAGERCEARLCP 356
 DB 404 YTGPAQSQDVDECALGAN-----PCBHAGKCLNTLISFEQCLQGTGPRCEIDV-- 453
 QY 357 EGIYGIKCKRC--PCHLENTHSCHPMSE--CAKPRMSGLYC----- 396
 DB 454 -----NECISNPC--QNDATCLDQIGEFQICLMPYEGYVCININDECASSPCLHN 502
 QY 397 -----NE--TCSPFYGEACQ--QICS--CONGADC--DSYTGKCTCAPRGKID 438
 DB 503 GRGVOKINFLQCCPRGFGHLCQYVDDECASTPCKNKAKCLDPRITTYCYCEGTGTH 562
 QY 439 CST-----PCPLATYINCSSRCGCKNDAYCSPVDSCTCKAGW----- 477
 DB 563 CEVDIDECDDPDCHIGL-----CK-DGVAT--FTCLCQDPYGTGHHCETINNECH 608
 QY 478 -----HGVDCSTR-----CPSGTWFGCINLTCQ-----CLNGACNTLDG-TCTCAP 518
 DB 609 SOPCRHGTGQCDNNTYLCCLKLGTTGPNCEINLDDCASNPDCDSGLDKIDGYECACBP 668
 QY 519 GMRGKCEL-----PCDDGTGVLNCAERCDSCSHADGCHPTT----- 554
 DB 669 GTTGSNCNNNIDECAGSPCHNGTCEBDJAGFTC-RC--PEGHDTCLSEVNECHS 722
 QY 555 -----GHCR-----CLPGNSGVHD-----SYCABG 575
 DB 723 NPCIHGACRDGLNGYKCDAPGMSGTNCINNNECESNPCCVNGTCKDMTSGYVTCRBS 782
 QY 576 RMGPNC-----SLPCY-----CKNKA 591
 DB 783 FSGPQNTININECASNPCLNCGTCLDIDVAGYKNCNPLPYTGATCEVYLAPCATSPCKNSG 842
 QY 592 SCSPDDGI-----CECAPGRGTTCQ-----RISCPFYH 622
 DB 843 VKRESDDYSPFCVPTGQGTCEIDINECYKSPCRHASCNTGSTRCLCQAGTGR 902
 QY 623 RSCQTCPOCVHSSGCPCH-----ITGLCDCLPFGTALCNE----- 658
 DB 903 NCSIDIDDC--RNPCHNGSGCTDGVNAFCOLPFGCAFCEDEDINECATNPCCQNGANC 960
 QY 659 -----VCPBGRGKACAG--ICT--CTNNGTCNPIDR--SCQCTPFGWIGSDC 699
 DB 961 TDCVDSYTCCTPTGFGNIGHCENNTPDCTESSCFNGTGC--VDGINSFTCLCPGFGTGSYC 1018

QY 700 -----SOP-----CPPAHMGPNC---IHTCN---CHNGAFRC----- 724
 DB 1019 QYDVNCEBDRPCLHGTGTCODSYGTGYKCTCPGGTGLNCLNLYRWCDSACKNGCKMQTN 1078
 QY 725 SAYDECKCTPMTGLYC--TORCPLGFY--GRDICALCQ----- 760
 DB 1079 TQY--HCECRSGWTGNCVDLVSVCEVAOKRQIDVTLTLCQHGGLCVDEEDKHYCHQAG 1136
 QY 761 -----CONGADC-DHISG-OCTCTGTGMRHCEBK----- 788
 DB 1137 YTGSYCEDEVDKCSNPCCNGATCTDYLGFSCKVAGYHGSNCSEINECLSQPCQNG 1196
 QY 789 -----CPSGTYGCGRQICD-----CLNNGSTC-DHITG--TCYCS 820
 DB 1197 TCIDLTNTYKKGSCPRGTQVHCEINVDCHPPLDPAKSPKCFNNCTCYDQVGGYTCTCP 1256
 QY 821 PGMKARCDQAGVITVNLN 840
 DB 1257 PGFVGRCR-----GDVN 1269

RESULT 5

NC2_HUMAN
 ID NC2_HUMAN STANDARD; PRT; 2471 AA.
 AC 004721; Q9H240; Q99734;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuregulin locus notch homolog protein 2 precursor (Notch 2) (hN2).
 GN NOTCH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RC TISSUE=Brain;
 RA Blaumüller C.M., Mann R.S.;
 RT "Complete human notch 2 (hN2) cDNA sequence.";
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RC TISSUE=FROM N.A.
 RA Cornea R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
 RT "Human Notch2, a novel member of cell-fate determining NOTCH
 family.";
 RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RC TISSUE=T-cell;
 RA Lemasson I., Devaux C., Mesnard J.M.;
 RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
 RL Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RC TISSUE=Brain;
 RA MEDLINE=93265135; PubMed=1303260;
 RA Stifani S., Blaumüller C.M., Redhead N.J., Hill R.E.,
 RT "Human homologs of a Drosophila enhancer of split gene product define
 a novel family of nuclear proteins.";
 RL Nat. Genet. 2:119-127(1992).
 RN [5]
 RC POST-TRANSLATIONAL PROCESSING.
 RA MEDLINE=97386453; PubMed=9244302;
 RA Blaumüller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
 RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on
 the plasma membrane.";
 RL Cell 90:281-291(1997).
 RN [6]
 RC IDENTIFICATION OF LIGANDS.
 RA MEDLINE=99180765; PubMed=10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,

RA Banks A., Lelman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands.
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, heart, kidney, lung,
 CC skeletal muscle and liver.
 CC -1- PPM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sb.ch).
 CC -----
 DR EMBL; AF308601; AAA36377.2; -;
 DR EMBL; AF315356; ANG37073.1; -;
 DR EMBL; U77493; AAB19224.1; -;
 DR HSSP; P00740; IEDM.
 DR Genev; HGNC:7882; NOTCH2.
 DR MIM; 600275; -;
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; ASX_hydroxyl.
 DR InterPro: IPR000361; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF_CA.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR002049; Lamlnln_EGF.
 DR InterPro: IPR000800; Notch.
 DR Pfam; PF00008; EGF; 35.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PRO0010; EGFBL00.
 DR PRINTS; PRO0011; EGFAMININ.
 DR SMART; SM00248; ANK; 4.
 DR SMART; SM00179; EGF_CA; 22.
 DR SMART; SM00001; EGF_Like; 12.
 DR SMART; SM00004; NL; 2.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF-1; 34.
 DR PROSITE; PS01186; EGF-2; 29.
 DR PROSITE; PS01187; EGF_CA; 22.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;

 KW Transmembrane; Glycoprotein; Signal; Phosphorylation.
 FT SIGNAL 1 25
 FT CHAIN 26 2471
 FT CHAIN 1666 2471
 FT CHAIN 1697 2471
 FT CHAIN 1677 2471
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2471
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 160
 FT DOMAIN 182 219
 FT DOMAIN 221 258
 FT DOMAIN 260 296
 FT DOMAIN 298 336
 FT DOMAIN 338 374
 FT DOMAIN 375 413
 FT DOMAIN 415 454
 FT DOMAIN 456 494
 FT DOMAIN 494 530
 FT DOMAIN 532 568
 FT DOMAIN 570 605
 FT DOMAIN 607 643
 FT DOMAIN 645 680
 FT DOMAIN 682 718
 FT DOMAIN 720 755
 FT DOMAIN 757 793
 FT DOMAIN 795 831
 FT DOMAIN 833 871
 FT DOMAIN 873 909
 FT DOMAIN 911 947
 FT DOMAIN 949 985
 FT DOMAIN 987 1023
 FT DOMAIN 1025 1061
 FT DOMAIN 1063 1099
 FT DOMAIN 1101 1147
 FT DOMAIN 1149 1185
 FT DOMAIN 1187 1223
 FT DOMAIN 1225 1262
 FT DOMAIN 1264 1302
 FT DOMAIN 1304 1343
 FT DOMAIN 1374 1412
 FT DOMAIN 1420 1456
 FT REPEAT 1456 1503
 FT REPEAT 1503 1535
 FT REPEAT 1535 1571
 FT REPEAT 1571 1607
 FT REPEAT 1607 1643
 FT REPEAT 1643 1679
 FT REPEAT 1679 1715
 FT REPEAT 1715 1751
 FT REPEAT 1751 1787
 FT REPEAT 1787 1823
 FT REPEAT 1823 1859
 FT REPEAT 1859 1895
 FT REPEAT 1895 1931
 FT REPEAT 1931 1967
 FT REPEAT 1967 2003
 FT REPEAT 2003 2038
 FT DOMAIN 1645 1681
 FT DOMAIN 1681 1717
 FT DOMAIN 1717 1753
 FT DOMAIN 1753 1789
 FT DOMAIN 1789 1825
 FT DOMAIN 1825 1861
 FT DOMAIN 1861 1897
 FT DOMAIN 1897 1933
 FT DOMAIN 1933 1969
 FT DOMAIN 1969 2005
 FT DOMAIN 2005 2041
 FT DOMAIN 2041 2077
 FT DOMAIN 2077 2113
 FT DOMAIN 2113 2149
 FT DOMAIN 2149 2185
 FT DOMAIN 2185 2221
 FT DOMAIN 2221 2257
 FT DOMAIN 2257 2293
 FT DOMAIN 2293 2329
 FT DOMAIN 2329 2365
 FT DOMAIN 2365 2401
 FT DOMAIN 2401 2437
 FT DOMAIN 2437 2473
 FT DOMAIN 2473 2509
 FT DOMAIN 2509 2545
 FT DOMAIN 2545 2581
 FT DOMAIN 2581 2617
 FT DOMAIN 2617 2653
 FT DOMAIN 2653 2689
 FT DOMAIN 2689 2725
 FT DOMAIN 2725 2761
 FT DOMAIN 2761 2797
 FT DOMAIN 2797 2833
 FT DOMAIN 2833 2869
 FT DOMAIN 2869 2905
 FT DOMAIN 2905 2941
 FT DOMAIN 2941 2977
 FT DOMAIN 2977 3013
 FT DOMAIN 3013 3049
 FT DOMAIN 3049 3085
 FT DOMAIN 3085 3121
 FT DOMAIN 3121 3157
 FT DOMAIN 3157 3193
 FT DOMAIN 3193 3229
 FT DOMAIN 3229 3265
 FT DOMAIN 3265 3301
 FT DOMAIN 3301 3337
 FT DOMAIN 3337 3373
 FT DOMAIN 3373 3409
 FT DOMAIN 3409 3445
 FT DOMAIN 3445 3481
 FT DOMAIN 3481 3517
 FT DOMAIN 3517 3553
 FT DOMAIN 3553 3589
 FT DOMAIN 3589 3625
 FT DOMAIN 3625 3661
 FT DOMAIN 3661 3697
 FT DOMAIN 3697 3733
 FT DOMAIN 3733 3769
 FT DOMAIN 3769 3805
 FT DOMAIN 3805 3841
 FT DOMAIN 3841 3877
 FT DOMAIN 3877 3913
 FT DOMAIN 3913 3949
 FT DOMAIN 3949 3985
 FT DOMAIN 3985 4021
 FT DOMAIN 4021 4057
 FT DOMAIN 4057 4093
 FT DOMAIN 4093 4129
 FT DOMAIN 4129 4165
 FT DOMAIN 4165 4201
 FT DOMAIN 4201 4237
 FT DOMAIN 4237 4273
 FT DOMAIN 4273 4309
 FT DOMAIN 4309 4345
 FT DOMAIN 4345 4381
 FT DOMAIN 4381 4417
 FT DOMAIN 4417 4453
 FT DOMAIN 4453 4489
 FT DOMAIN 4489 4525
 FT DOMAIN 4525 4561
 FT DOMAIN 4561 4597
 FT DOMAIN 4597 4633
 FT DOMAIN 4633 4669
 FT DOMAIN 4669 4705
 FT DOMAIN 4705 4741
 FT DOMAIN 4741 4777
 FT DOMAIN 4777 4813
 FT DOMAIN 4813 4849
 FT DOMAIN 4849 4885
 FT DOMAIN 4885 4921
 FT DOMAIN 4921 4957
 FT DOMAIN 4957 4993
 FT DOMAIN 4993 5029
 FT DOMAIN 5029 5065
 FT DOMAIN 5065 5101
 FT DOMAIN 5101 5137
 FT DOMAIN 5137 5173
 FT DOMAIN 5173 5209
 FT DOMAIN 5209 5245
 FT DOMAIN 5245 5281
 FT DOMAIN 5281 5317
 FT DOMAIN 5317 5353
 FT DOMAIN 5353 5389
 FT DOMAIN 5389 5425
 FT DOMAIN 5425 5461
 FT DOMAIN 5461 5497
 FT DOMAIN 5497 5533
 FT DOMAIN 5533 5569
 FT DOMAIN 5569 5605
 FT DOMAIN 5605 5641
 FT DOMAIN 5641 5677
 FT DOMAIN 5677 5713
 FT DOMAIN 5713 5749
 FT DOMAIN 5749 5785
 FT DOMAIN 5785 5821
 FT DOMAIN 5821 5857
 FT DOMAIN 5857 5893
 FT DOMAIN 5893 5929
 FT DOMAIN 5929 5965
 FT DOMAIN 5965 6001
 FT DOMAIN 6001 6037
 FT DOMAIN 6037 6073
 FT DOMAIN 6073 6109
 FT DOMAIN 6109 6145
 FT DOMAIN 6145 6181
 FT DOMAIN 6181 6217
 FT DOMAIN 6217 6253
 FT DOMAIN 6253 6289
 FT DOMAIN 6289 6325
 FT DOMAIN 6325 6361
 FT DOMAIN 6361 6397
 FT DOMAIN 6397 6433
 FT DOMAIN 6433 6469
 FT DOMAIN 6469 6505
 FT DOMAIN 6505 6541
 FT DOMAIN 6541 6577
 FT DOMAIN 6577 6613
 FT DOMAIN 6613 6649
 FT DOMAIN 6649 6685
 FT DOMAIN 6685 6721
 FT DOMAIN 6721 6757
 FT DOMAIN 6757 6793
 FT DOMAIN 6793 6829
 FT DOMAIN 6829 6865
 FT DOMAIN 6865 6901
 FT DOMAIN 6901 6937
 FT DOMAIN 6937 6973
 FT DOMAIN 6973 7009
 FT DOMAIN 7009 7045
 FT DOMAIN 7045 7081
 FT DOMAIN 7081 7117
 FT DOMAIN 7117 7153
 FT DOMAIN 7153 7189
 FT DOMAIN 7189 7225
 FT DOMAIN 7225 7261
 FT DOMAIN 7261 7297
 FT DOMAIN 7297 7333
 FT DOMAIN 7333 7369
 FT DOMAIN 7369 7405
 FT DOMAIN 7405 7441
 FT DOMAIN 7441 7477
 FT DOMAIN 7477 7513
 FT DOMAIN 7513 7549
 FT DOMAIN 7549 7585
 FT DOMAIN 7585 7621
 FT DOMAIN 7621 7657
 FT DOMAIN 7657 7693
 FT DOMAIN 7693 7729
 FT DOMAIN 7729 7765
 FT DOMAIN 7765 7801
 FT DOMAIN 7801 7837
 FT DOMAIN 7837 7873
 FT DOMAIN 7873 7909
 FT DOMAIN 7909 7945
 FT DOMAIN 7945 7981
 FT DOMAIN 7981 8017
 FT DOMAIN 8017 8053
 FT DOMAIN 8053 8089
 FT DOMAIN 8089 8125
 FT DOMAIN 8125 8161
 FT DOMAIN 8161 8197
 FT DOMAIN 8197 8233
 FT DOMAIN 8233 8269
 FT DOMAIN 8269 8305
 FT DOMAIN 8305 8341
 FT DOMAIN 8341 8377
 FT DOMAIN 8377 8413
 FT DOMAIN 8413 8449
 FT DOMAIN 8449 8485
 FT DOMAIN 8485 8521
 FT DOMAIN 8521 8557
 FT DOMAIN 8557 8593
 FT DOMAIN 8593 8629
 FT DOMAIN 8629 8665
 FT DOMAIN 8665 8701
 FT DOMAIN 8701 8737
 FT DOMAIN 8737 8773
 FT DOMAIN 8773 8809
 FT DOMAIN 8809 8845
 FT DOMAIN 8845 8881
 FT DOMAIN 8881 8917
 FT DOMAIN 8917 8953
 FT DOMAIN 8953 8989
 FT DOMAIN 8989 9025
 FT DOMAIN 9025 9061
 FT DOMAIN 9061 9097
 FT DOMAIN 9097 9133
 FT DOMAIN 9133 9169
 FT DOMAIN 9169 9205
 FT DOMAIN 9205 9241
 FT DOMAIN 9241 9277
 FT DOMAIN 9277 9313
 FT DOMAIN 9313 9349
 FT DOMAIN 9349 9385
 FT DOMAIN 9385 9421
 FT DOMAIN 9421 9457
 FT DOMAIN 9457 9493
 FT DOMAIN 9493 9529
 FT DOMAIN 9529 9565
 FT DOMAIN 9565 9601
 FT DOMAIN 9601 9637
 FT DOMAIN 9637 9673
 FT DOMAIN 9673 9709
 FT DOMAIN 9709 9745
 FT DOMAIN 9745 9781
 FT DOMAIN 9781 9817
 FT DOMAIN 9817 9853
 FT DOMAIN 9853 9889
 FT DOMAIN 9889 9925
 FT DOMAIN 9925 9961
 FT DOMAIN 9961 9997
 FT DOMAIN 9997 10033
 FT DOMAIN 10033 10069
 FT DOMAIN 10069 10105
 FT DOMAIN 10105 10141
 FT DOMAIN 10141 10177
 FT DOMAIN 10177 10213
 FT DOMAIN 10213 10249
 FT DOMAIN 10249 10285
 FT DOMAIN 10285 10321
 FT DOMAIN 10321 10357
 FT DOMAIN 10357 10393
 FT DOMAIN 10393 10429
 FT DOMAIN 10429 10465
 FT DOMAIN 10465 10501
 FT DOMAIN 10501 10537
 FT DOMAIN 10537 10573
 FT DOMAIN 10573 10609
 FT DOMAIN 10609 10645
 FT DOMAIN 10645 10681
 FT DOMAIN 10681 10717
 FT DOMAIN 10717 10753
 FT DOMAIN 10753 10789
 FT DOMAIN 10789 10825
 FT DOMAIN 10825 10861
 FT DOMAIN 10861 10897
 FT DOMAIN 10897 10933
 FT DOMAIN 10933 10969
 FT DOMAIN 10969 11005
 FT DOMAIN 11005 11041
 FT DOMAIN 11041 11077
 FT DOMAIN 11077 11113
 FT DOMAIN 11113 11149
 FT DOMAIN 11149 11185
 FT DOMAIN 11185 11221
 FT DOMAIN 11221 11257
 FT DOMAIN 11257 11293
 FT DOMAIN 11293 11329
 FT DOMAIN 11329 11365
 FT DOMAIN 11365 11401
 FT DOMAIN 11401 11437
 FT DOMAIN 11437 11473
 FT DOMAIN 11473 11509
 FT DOMAIN 11509 11545
 FT DOMAIN 11545 11581
 FT DOMAIN 11581 11617
 FT DOMAIN 11617 11653
 FT DOMAIN 11653 11689
 FT DOMAIN 11689 11725
 FT DOMAIN 11725 11761
 FT DOMAIN 11761 11797
 FT DOMAIN 11797 11833
 FT DOMAIN 11833 11869
 FT DOMAIN 11869 11905
 FT DOMAIN 11905 11941
 FT DOMAIN 11941 11977
 FT DOMAIN 11977 12013
 FT DOMAIN 12013 12049
 FT DOMAIN 12049 12085
 FT DOMAIN 12085 12121
 FT DOMAIN 12121 12157
 FT DOMAIN 12157 12193
 FT DOMAIN 12193 12229
 FT DOMAIN 12229 12265
 FT DOMAIN 12265 12301
 FT DOMAIN 12301 12337
 FT DOMAIN 12337 12373
 FT DOMAIN 12373 12409
 FT DOMAIN 12409 12445
 FT DOMAIN 12445 12481
 FT DOMAIN 12481 12517
 FT DOMAIN 12517 12553
 FT DOMAIN 12553 12589
 FT DOMAIN 12589 12625
 FT DOMAIN 12625 12661
 FT DOMAIN 12661 12697
 FT DOMAIN 12697 12733
 FT DOMAIN 12733 12769
 FT DOMAIN 12769 12805
 FT DOMAIN 12805 12841
 FT DOMAIN 12841 12877
 FT DOMAIN 12877 12913
 FT DOMAIN 12913 12949
 FT DOMAIN 12949 12985
 FT DOMAIN 12985 13021
 FT DOMAIN 13021 13057
 FT DOMAIN 13057 13093
 FT DOMAIN 13093 13129
 FT DOMAIN 13129 13165
 FT DOMAIN 13165 13201
 FT DOMAIN 13201 13237
 FT DOMAIN 13237 13273
 FT DOMAIN 13273 13309
 FT DOMAIN 13309 13345
 FT DOMAIN 13345 13381
 FT DOMAIN 13381 13417
 FT DOMAIN 13417 13453
 FT DOMAIN 13453 13489
 FT DOMAIN 13489 13525
 FT DOMAIN 13525 13561
 FT DOMAIN 13561 13597
 FT DOMAIN 13597 13633
 FT DOMAIN 13633 13669
 FT DOMAIN 13669 13705
 FT DOMAIN 13705 13741
 FT DOMAIN 13741 13777
 FT DOMAIN 13777 13813
 FT DOMAIN 13813 13849
 FT DOMAIN 13849 13885
 FT DOMAIN 13885 13921
 FT DOMAIN 13921 13957
 FT DOMAIN 13957 14000
 FT DOMAIN 14000 14043
 FT DOMAIN 14043 14086
 FT DOMAIN 14086 14129
 FT DOMAIN 14129 14172
 FT DOMAIN 14172 14215
 FT DOMAIN 14215 14258
 FT DOMAIN 14258 14301
 FT DOMAIN 14301 14344
 FT DOMAIN 14344 14387
 FT DOMAIN 14387 14430
 FT DOMAIN 14430 14473
 FT DOMAIN 14473 14516
 FT DOMAIN 14516 14559
 FT DOMAIN 14559 14602
 FT DOMAIN 14602 14645
 FT DOMAIN 14645 14688
 FT DOMAIN 14688 14731
 FT DOMAIN 14731 14774
 FT DOMAIN 14774 14817
 FT DOMAIN 14817 14860
 FT DOMAIN 14860 14903
 FT DOMAIN 14903 14946
 FT DOMAIN 14946 14989
 FT DOMAIN 14989 15032
 FT DOMAIN 15032 15075
 FT DOMAIN 15075 15118
 FT DOMAIN 15118 15161
 FT DOMAIN 15161 15204
 FT DOMAIN 15204 15247
 FT DOMAIN 15247 15290
 FT DOMAIN 15290 15333
 FT DOMAIN 15333 15376
 FT DOMAIN 15376 15419
 FT DOMAIN 15419 15462
 FT DOMAIN 15462 15505
 FT DOMAIN 15505 15548
 FT DOMAIN 15548 15591
 FT DOMAIN 15591 15634
 FT DOMAIN 15634 15677
 FT DOMAIN 15677 15720
 FT DOMAIN 15720 15763
 FT DOMAIN 15763 15806
 FT DOMAIN 15806 15849
 FT DOMAIN 15849 15892
 FT DOMAIN 15892 15935
 FT DOMAIN 15935 15978
 FT DOMAIN 15978 16021
 FT DOMAIN 16021 16064
 FT DOMAIN 16064 16107
 FT DOMAIN 16107 16150
 FT DOMAIN 16150 16193
 FT DOMAIN 16193 16236
 FT DOMAIN 16236 16279
 FT DOMAIN 16279 16322
 FT DOMAIN 16322 16365
 FT DOMAIN 16365 16408
 FT DOMAIN 16408 16451
 FT DOMAIN 16451 16494
 FT DOMAIN 16494 16537
 FT DOMAIN 16537 16580
 FT DOMAIN 16580 16623
 FT DOMAIN 16623 16666
 FT DOMAIN 16666 16709
 FT DOMAIN 16709 16752
 FT DOMAIN 16752 16795
 FT DOMAIN 16795 16838
 FT DOMAIN 16838 16881
 FT DOMAIN 16881 16924
 FT DOMAIN 16924 16967
 FT DOMAIN 16967 17010
 FT DOMAIN 17010 17053
 FT DOMAIN 17053 17096
 FT DOMAIN 17096 17139
 FT DOMAIN 17139 17182
 FT DOMAIN 17182 17225
 FT DOMAIN 17225 17268
 FT DOMAIN 17268 17311
 FT DOMAIN 17311 17354
 FT DOMAIN 17354 17397
 FT DOMAIN 17397 17440
 FT DOMAIN 17440 17483
 FT DOMAIN 17483 17526
 FT DOMAIN 17526 17569
 FT DOMAIN 17569 17612
 FT DOMAIN 17612 17655
 FT DOMAIN 17655 17698
 FT DOMAIN 17698 17741
 FT DOMAIN 17741 17784
 FT DOMAIN 17784 17827
 FT DOMAIN 17827 17870
 FT DOMAIN 17870 17913
 FT DOMAIN 17913 17956
 FT DOMAIN 17956 17999
 FT DOMAIN 17999 18042
 FT DOMAIN 18042 18085
 FT DOMAIN 18085 18128
 FT DOMAIN 18128 18171
 FT DOMAIN 18171 18214
 FT DOMAIN 18214 18257
 FT DOMAIN 18257 18300
 FT DOMAIN 18300 18343
 FT DOMAIN 18343 18386
 FT DOMAIN 18386 18429
 FT DOMAIN 18429 18472
 FT DOMAIN 18472 18515
 FT DOMAIN 18515 18558
 FT DOMAIN 18558 18601
 FT DOMAIN 18601 18644
 FT DOMAIN 18644 18687
 FT DOMAIN 18687 18730
 FT DOMAIN 18730 18773
 FT DOMAIN 18773 18816
 FT DOMAIN 18816 18859
 FT DOMAIN 18859 18902
 FT DOMAIN 18902 18945
 FT DOMAIN 18945 18988
 FT DOMAIN 18988 19031
 FT DOMAIN 19031 19074
 FT DOMAIN 19074 19117
 FT DOMAIN 19117 19160
 FT DOMAIN 19160 19203
 FT DOMAIN 19203 19246
 FT DOMAIN 19246 19289
 FT DOMAIN 19289 19332
 FT DOMAIN 19332 19375
 FT DOMAIN 19375 19418
 FT DOMAIN 19418 19461
 FT DOMAIN 19461 19504
 FT DOMAIN 19504 19547
 FT DOMAIN 19547 19590
 FT DOMAIN 19590 19633
 FT DOMAIN 19633 19676
 FT DOMAIN 19676 19719
 FT DOMAIN 19719 19762
 FT DOMAIN 19762 19805
 FT DOMAIN 19805 19848
 FT DOMAIN 19848 19891
 FT DOMAIN 19891 19934
 FT DOMAIN 19934 19977
 FT DOMAIN 19977 20020
 FT DOMAIN 20020 20063
 FT DOMAIN 20063 20106
 FT DOMAIN 20106 20149
 FT DOMAIN 20149 20192
 FT DOMAIN 20192 20235
 FT DOMAIN 20235 20278
 FT DOMAIN 20278 20321
 FT DOMAIN 20321 20364
 FT DOMAIN 20364 20407
 FT DOMAIN 20407 20450
 FT DOMAIN 20450 20493
 FT DOMAIN 20493 20536
 FT DOMAIN 20536 20579
 FT DOMAIN 20579 20622
 FT DOMAIN 20622 20665
 FT DOMAIN 20665 20708
 FT DOMAIN 20708 20751
 FT DOMAIN 20751 20794
 FT DOMAIN 20794 20837
 FT DOMAIN 20837 20880
 FT DOMAIN 20880 20923
 FT DOMAIN 20923 20966
 FT DOMAIN 20966 21009
 FT DOMAIN 21009 21052
 FT DOMAIN 21052 21095
 FT DOMAIN 21095 21138
 FT DOMAIN 21138 21181
 FT DOMAIN 21181 21224
 FT DOMAIN 21224 21267
 FT DOMAIN 21267 21310
 FT DOMAIN 21310 21353
 FT DOMAIN 21353 21396
 FT DOMAIN 21396 21439
 FT DOMAIN 21439 21482
 FT DOMAIN 21482 21525
 FT DOMAIN 21525 21568
 FT DOMAIN 21568 21611
 FT DOMAIN 21611 21654
 FT DOMAIN 21654 21697
 FT DOMAIN 21697 21740
 FT DOMAIN 21740 21783
 FT DOMAIN 21783 21826
 FT DOMAIN 21826 21869
 FT DOMAIN 21869 21912
 FT DOMAIN 21912 21955
 FT DOMAIN 21955 21998
 FT DOMAIN 21998 22041
 FT DOMAIN 22041 22084
 FT DOMAIN 22084 22127
 FT DOMAIN 22127 22170
 FT DOMAIN 22170 22213
 FT DOMAIN 22213 22256
 FT DOMAIN 22256 22299
 FT DOMAIN 22299 22342
 FT DOMAIN 22342 22385
 FT DOMAIN 22385 22428
 FT DOMAIN 22428 22471
 FT DOMAIN 22471 22514
 FT DOMAIN 22514 22557
 FT DOMAIN 22557 22600
 FT DOMAIN 22600 22643
 FT DOMAIN 22643 22686
 FT DOMAIN 22686 22729
 FT DOMAIN 22729 22772
 FT DOMAIN 22772 22815
 FT DOMAIN 22815 22858
 FT DOMAIN 22858 22901
 FT DOMAIN 22901 22944
 FT DOMAIN 22944 22987
 FT DOMAIN 22987 23030
 FT DOMAIN 23030 23073
 FT DOMAIN 23073 23116
 FT DOMAIN 23116 23159
 FT DOMAIN 23159 23202
 FT DOMAIN 23202 23245
 FT DOMAIN 23245 23288
 FT DOMAIN 23288 23331
 FT DOMAIN 23331 23374
 FT DOMAIN 23374 23417
 FT DOMAIN 23417 23460
 FT DOMAIN 23460 23503
 FT DOMAIN 23503 23546
 FT DOMAIN 23546 23589
 FT DOMAIN 23589 23632
 FT DOMAIN 23632 23675
 FT DOMAIN 23675 23

Query Match	14.8%;	Score 998;	DB 1;	Length 2471;
Best Local Similarity	-24.4%;	Pred. No. 4.8e-51;		
Matches 321;	Conservative . 79;	Mismatches 322;	Indels 596;	Gaps 70

QY	465	-----SP-VDGSGT-----	-----CAAGHHGADCSIT-----	484		
Db	714	HPSCSYQVNECLSSPCLHGHGCTGGISGYKCLDAGVNGVINCEDYKNECLSNPCQNGTGN	773			
QY	485	-----RCPSGTWGFGCNLTQO-----	-----CLNGGAC-----	507		
Db	774	NLVNGYKCTCKKGFKNQCNVNIIDECASNPCLNQGTCILDVSYGTYCHMLPYTGAKNQCY	833			
QY	508	-----NTLDGTCAPARNGECCELPCODGTYGLCAERCCSHAD	548			
Db	834	LAPSPNPCEMAAVCKEAPNEFESFTCLCAPMGOGKCTVDVE-----	CVSK-PCMNG	886		
QY	549	GCHPTTGH-----CRLPGMSYGVHCDYVCAEGWMPNCSLPCYCKNGASC--	SPDGTICEAP	604		
Db	887	ICHHTQSYSTCECPGFGSGHDEE-----	DINDCLANPCQNGSGCDKNTVFSCILP	939		
QY	605	GFRGTCOR-----	-----ICSPGFYGRHSQTCPOCVHSS-----	635		
Db	940	GFVGDCKQDTMNBELSPCKNGSTCDYVNSYTCCTPAFPHGVHCENNIIDECTESSFNG	999			
QY	636	GRCHHTTGL-----CDCLPGFTGALC-----	NE-----	VCPSGRFG	666	
Db	1000	GTC--VDGINSFSCLCPVGFPTGPFCLHDINECSSNPCLNSGTCVVDGLGTYRCTCPLGYTG	1057			
QY	667	KNC--AGICT-----CYNNGC--	NPIDRSQCQPGMIGSDCSQ-----	701		
Db	1058	KNCQTLVNLCSPSPCFKKNGCAOEKARPRCLCPGMDGAYCDVLNVSCAALAQKVPVE	1117			
QY	702	-----PCPPAHMGPNC-----	-----IHTC-----	NCHNGAFCSAYDG--	BCKCP	735
Db	1118	HLCHSGSITCINAGNTHHCQCPPLGYTSGYCEBQLEDCASNPQCHGATCSDFIGGYKCEVP	1177			
QY	736	GMTGLYCTOR-----	-----CPIGFTG-----	KDCALICQCON	763	
Db	1178	GYGVNCEYEVWDECONQPCQNGSTCIDLVNHFKSCPPGFRGLICEBENIDCAGAPHCLN	1237			
QY	764	GADC-DHISG-----	-----	OCRTGFM	781	
Db	1238	GGQCVDRIGYSGCRCLPFGAERCEBGDINECLSNPCSSBSGLDCLDKNNYCCVCRSAFT	1297			
QY	782	GRHCE--OKPCSGTYGYGRCRLCDCLNNSYCDHNGT-----	-----CYCSPGMKARGAQDQ	831		
Db	1298	GRHCETPLDYCPK-----	-----PCLNGSTCAVANSVWDGFLCRCPPFGSARGQSS	1345		

RESULT 7

NTC2_MOUSE STANDARD: PRT: 2470 AA.

AC 035516; 060941; 006008;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch B).

GN NOTCH2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

ON NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA STRAIN=C57BL/6; TISSUE=Thymus;

RC Hamada Y., Higuchi M., Tsujimoto Y.;

RT "Complete amino acid sequence and multiform transcripts encoded by a single copy of mouse Notch2 gene."

RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 316-1518 FROM N.A.

RA STRAIN=C57BL/6 X CBA; TISSUE=Embryo;

RC MEDLINE=93178563; PubMed=8440332;

RA Lardelli M., Lendahl U.;

RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a wide variety of tissues."

RL Exp. Cell Res. 204:364-372(1993).
 [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RX MEDLINE-97075110; PubMed-8917536;
 RA Milner L.A., Bigas A., Kopan R., Braehem-Stein C., Bernstein I.D.,
 RT Martin D.I.;
 "Inhibition of granulocytic differentiation by mNotch1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN [4]
 RP FUNCTION.
 RX MEDLINE-99396706; PubMed-10393120;
 RA Hamada Y., Kadohawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 embryonic lethality.";
 RT Development 126:3415-3424(1999).
 RN [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RX MEDLINE-95333893; PubMed-7609614;
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 mouse brain.";
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE-21523956; PubMed-11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent
 proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RX MEDLINE-21374376; PubMed-11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 Upon ligand activation through the released notch intracellular
 domain (NICD) it forms a transcriptional activator complex with
 RBP-J kappa and activates genes of the enhancer of split locus.
 Affects the implementation of differentiation, proliferation and
 apoptotic programs (By similarity). May play an essential role in
 postimplantation development, probably in some aspect of cell
 specification and/or differentiation.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 terminal fragment N(EC) which are probably linked by disulfide
 bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
 produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 neuroepithelia, somites, optic vesicles and branchial arches, but
 not heart.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 the postnatal endodermal cells, and the choroid plexus throughout
 embryonic and postnatal development.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 which is proteolytically cleaved by a furin-like convertase in the
 trans-Golgi network before it reaches the plasma membrane to yield
 an active, ligand-accessible form. Cleavage results in a C-
 terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 ligand binding, it is cleaved by TNF-alpha converting enzyme
 (TACE) to yield a membrane-associated intermediate fragment called
 notch extracellular truncation (NEXT). This fragment is then
 cleaved by presenilin dependent gamma-secretase to release a
 notch-derived peptide containing the intracellular domain (NICD)
 from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (see <http://www.isb-sb.ch/announce/>
 CC or send an email to license@isb-sb.ch).
 CC -----
 CC EMBL: D32210; BAA22094.1; -
 CC EMBL: X68279; CA448340.1; -
 CC EMBL: U31881; AAC52924.1; -
 CC HSSP: P61009; IFSB.
 CC MGD: MGI:97364; Notch2.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_II.
 CC InterPro: IPR000800; Notch.
 CC Pfam: PF00008; EGF_35.
 CC Pfam: PF00023; ank; 6.
 CC Pfam: PF00066; notch; 2.
 CC PRINTS: PR00010; EGFLOOD.
 CC PRINTS: PR01452; NOTCH.
 CC SMART: SM00248; ANK; 4.
 CC SMART: SM00179; EGF_CA; 22.
 CC SMART: SM00001; EGF-like; 12.
 CC SMART: SM00004; NL; 3.
 CC PROSITE: PS50088; ANK_REPEAT; 4.
 CC PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 22.
 CC PROSITE: PS00022; EGF_1; 33.
 CC PROSITE: PS01186; EGF_2; 27.
 CC PROSITE: PS01187; EGF_CA; 22.
 CC Receptor; Transcription regulation; Activator; Differentiation;
 CC development; protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing.
 CC KW
 CC SIGNAL 1 25
 CC FT CHAIN 26 2470
 CC FT CHAIN 1666 2470
 CC FT CHAIN 1667 2470
 CC FT CHAIN 1677 2470
 CC FT CHAIN 1678 2470
 CC FT TRANSMEM 1699 2470
 CC FT DOMAIN 26 63
 CC FT DOMAIN 64 102
 CC FT DOMAIN 105 143
 CC FT DOMAIN 144 180
 CC FT DOMAIN 182 219
 CC FT DOMAIN 221 256
 CC FT DOMAIN 258 294
 CC FT DOMAIN 296 332
 CC FT DOMAIN 336 374
 CC FT DOMAIN 373 411
 CC FT DOMAIN 413 452
 CC FT DOMAIN 454 490
 CC FT DOMAIN 492 528
 CC FT DOMAIN 530 566
 CC FT DOMAIN 568 603
 CC FT DOMAIN 605 641
 CC FT DOMAIN 643 678
 CC FT DOMAIN 680 716
 CC FT DOMAIN 718 753
 CC FT DOMAIN 755 791
 CC FT DOMAIN 793 829
 CC FT DOMAIN 831 869
 CC FT DOMAIN 871 907
 CC FT DOMAIN 909 945
 CC -----
 CC POTENTIAL.
 CC NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 CC NOTCH EXTRACELLULAR TRUNCATION.
 CC NOTCH INTRACELLULAR DOMAIN.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC EGF-LIKE 1.
 CC EGF-LIKE 2.
 CC EGF-LIKE 3.
 CC EGF-LIKE 4.
 CC EGF-LIKE 5.
 CC EGF-LIKE 6 (INCOMPLETE).
 CC EGF-LIKE 7.
 CC EGF-LIKE 8.
 CC EGF-LIKE 9.
 CC EGF-LIKE 10.
 CC EGF-LIKE 11.
 CC EGF-LIKE 12.
 CC EGF-LIKE 13.
 CC EGF-LIKE 14.
 CC EGF-LIKE 15.
 CC EGF-LIKE 16.
 CC EGF-LIKE 17.
 CC EGF-LIKE 18.
 CC EGF-LIKE 19.
 CC EGF-LIKE 20.
 CC EGF-LIKE 21.
 CC EGF-LIKE 22.
 CC EGF-LIKE 23.
 CC EGF-LIKE 24.
 CC EGF-LIKE 25.
 CC EGF-LIKE 26.
 CC EGF-LIKE 27.
 CC EGF-LIKE 28.
 CC EGF-LIKE 29.
 CC EGF-LIKE 30.
 CC EGF-LIKE 31.
 CC EGF-LIKE 32.
 CC EGF-LIKE 33.
 CC EGF-LIKE 34.
 CC EGF-LIKE 35.
 CC EGF-LIKE 36.
 CC EGF-LIKE 37.
 CC EGF-LIKE 38.
 CC EGF-LIKE 39.
 CC EGF-LIKE 40.
 CC EGF-LIKE 41.
 CC EGF-LIKE 42.
 CC EGF-LIKE 43.
 CC EGF-LIKE 44.
 CC EGF-LIKE 45.
 CC EGF-LIKE 46.
 CC EGF-LIKE 47.
 CC EGF-LIKE 48.
 CC EGF-LIKE 49.
 CC EGF-LIKE 50.
 CC EGF-LIKE 51.
 CC EGF-LIKE 52.
 CC EGF-LIKE 53.
 CC EGF-LIKE 54.
 CC EGF-LIKE 55.
 CC EGF-LIKE 56.
 CC EGF-LIKE 57.
 CC EGF-LIKE 58.
 CC EGF-LIKE 59.
 CC EGF-LIKE 60.
 CC EGF-LIKE 61.
 CC EGF-LIKE 62.
 CC EGF-LIKE 63.
 CC EGF-LIKE 64.
 CC EGF-LIKE 65.
 CC EGF-LIKE 66.
 CC EGF-LIKE 67.
 CC EGF-LIKE 68.
 CC EGF-LIKE 69.
 CC EGF-LIKE 70.
 CC EGF-LIKE 71.
 CC EGF-LIKE 72.
 CC EGF-LIKE 73.
 CC EGF-LIKE 74.
 CC EGF-LIKE 75.
 CC EGF-LIKE 76.
 CC EGF-LIKE 77.
 CC EGF-LIKE 78.
 CC EGF-LIKE 79.
 CC EGF-LIKE 80.
 CC EGF-LIKE 81.
 CC EGF-LIKE 82.
 CC EGF-LIKE 83.
 CC EGF-LIKE 84.
 CC EGF-LIKE 85.
 CC EGF-LIKE 86.
 CC EGF-LIKE 87.
 CC EGF-LIKE 88.
 CC EGF-LIKE 89.
 CC EGF-LIKE 90.
 CC EGF-LIKE 91.
 CC EGF-LIKE 92.
 CC EGF-LIKE 93.
 CC EGF-LIKE 94.
 CC EGF-LIKE 95.
 CC EGF-LIKE 96.
 CC EGF-LIKE 97.
 CC EGF-LIKE 98.
 CC EGF-LIKE 99.
 CC EGF-LIKE 100.

```

ET DOMAIN 947 983 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 985 1021 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1023 1059 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1061 1097 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1099 1145 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1147 1183 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1185 1221 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1223 1260 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1262 1300 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1302 1345 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL)
ET DOMAIN 1372 1410 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL)
ET REPEAT 1418 1454 LIN/NOTCH 1.
ET REPEAT 1501 1533 LIN/NOTCH 2.
ET REPEAT 1825 1869 ANK 1.
ET REPEAT 1874 1903 ANK 2.
ET REPEAT 1907 1937 ANK 3.
ET REPEAT 1941 1970 ANK 4.
ET REPEAT 1974 2003 ANK 5.
ET REPEAT 2007 2036 ANK 6.
ET DOMAIN 1645 1648 POLY-ALA.
ET DOMAIN 1992 1995 POLY-LEU.
ET DOMAIN 2183 2189 POLY-SER.
ET DOMAIN 2423 2428 POLY-GLY.
ET DOMAIN 2445 2450 POLY-GLY.
ET DISULFID 28 41 BY SIMILARITY.
ET DISULFID 35 51 BY SIMILARITY.
ET DISULFID 53 62 BY SIMILARITY.
ET DISULFID 79 79 BY SIMILARITY.

```

Query Match 14.7%; Score 993; DB 1; Length 2470;
 Best Local Similarity 25.8%; Pred. No. 9.5e-51;
 Matches 316; Conservative 87; Mismatches 295; Indels 526; Gaps 71;

```

QY 93 QCCGFESGMCYPRHADCYVHR-CIAPNTCO-----CEGWSGTNCSAAG 141
DB 91 RCAGCF--TGEDCYSTSHPCFYSRPONGGTCHMLSDYECYQVGFTRKQC-----142
QY 142 DHMPCSTRQCKNGALCNPTG--ACHCAFGWRNC-----DRCEQ-----184
DB 143 -QWIDALSL-PCENSGSTSVASQFSCKPAGLTGOKCEADINECDIPGKQNGGTCIN 200
QY 185 -GTGMDG----HQRQ-----CONCATDHY--TGECRPPGYTGAFC--D 223
DB 201 LPSGYRCQCGGFTYQHCDSBYVRLPCVNGGTCRGTGDFLECNCLPGEFSGYCERNID 260
QY 224 LCPRGK--HQPQCEQ-----RCP-----CONGVCHHTG-- 251
DB 261 DCPHNKCONGVGVGVNTYNCRCRPPMTGQFCTEDYDECLLPNACONGGTCTNRNGY 320
QY 252 ECSCPSGMGTVCQGP-----CPGRFGKNC--SQEC-- 281
DB 321 GCVCVNMSSGDDCEBNIDDCAYASCTPGSTCIDRVASFSLCPSGKAGLCHLDALISN 380
QY 282 QCHNGTCD--ATGQ--CHCSPGYTGERCO--DECPVGTGYLAETCCQVNGRCYH 334
DB 381 PCHGALCDTPPLNGOYICPCPGYKAGADCTEDVDECAM--ANSNPEHAHAKCVN 433
QY 335 VSQA--CLCEAGFGERCE-----ARLCEGLYGTCKDKR- 367
DB 434 TDGAFHCECLKGYAGPCREMDINECHSDPCONDATCLDKTIGFTCLCPGKGVHCELEV 493
QY 368 -----CP-----CHLE-----NTHSC--HPMSGE 384
DB 494 NEQSNPCVNNQCVDVYNNRQCICCPGFTGVPQOIIDDCSSITPLNGAKCIDHPNGYE 553
QY 385 CACPGMSGLYCNET-----CSPGYGAC--QOI-----412
DB 554 CQCATFTGILCDENINCPDPCPHGQCDGIDSYTCINPGYGAICSDQIDECYSSP 613
QY 413 -----GSCQNA-----DCDS-----VTG-----KTCAGPFG 435
DB 614 CLNDRCIDLVNGYQCNCPGTSGLNCINEFDCAISNPMHGVGVNDINRYSCVSPGFT 673

```

```

QY 436 GIDC-----SNPCPLGYTGNCSS--RCQCK-----NDAYCSP-VDSGCT 472
DB 674 GQRCNIDIDECASNPCKKATCINDVNGFRICPEBHPHSCYSQVNECSNPCTHNGCT 733
QY 473 -----CKAGHGVDCSI-----RPSGTWFGGNTL 498
DB 734 GGLSGYICLDDAGVGVNCEVDKNECLSNPQNGGTCNNLVNGYRCCKKGFKNQCVN 793
QY 499 CQ-----CLNGAC-----NTL 510
DB 794 IDECASNPCLNOGTCPFDVSGYTCCHMLPYTGKNCQTVLAPCSNPENNAVCKEADNFE 853
QY 511 DGTGTAPGRNGEXCELPCODGTGYGLNCAERDCSHADGCHPTGH--CRCLPMSGVHC 568
DB 854 SFSCLCAPGNQKRCCTDYDE-----CIK-PCMANNGVCHNTGSYVCECPFGSGMD 906
QY 569 DSVCAEGRWGPNCSLPCYCKRNGASCSPDDGI--CECAGFRGTTQOR-----613
DB 907 EE-----DINDLANPCONGGSCV--DHVNFSCCHGFGIDKQOTDMNECLSEPRK 957
QY 614 -----ICSPGYGHRCSQTCPCQVHSS--GPCNHTGL-----CDCLPFG 651
DB 958 NGTCSQVNSYTCCTCPAGFHVHCENNIDECTESSCFNGGTC--VDGINSFSLCPVGF 1015
QY 652 TGALC--NE-----VCPSGRFGKNC--AGICT--CTNNGT 680
DB 1016 TGPFCLDINDICSNPCLNAGTCVDGLGTYRCITPLGTGKNCQTVLNLCSRSCKKGT 1075
QY 681 C-NPIDRSCQYPRGWTGSDC--SQPCPPA--HMGPNCTHTNCNHNAGFCSAYD-GECK 732
DB 1076 CVGKARPHCLCPGMDGAVCDVLYNSCKAALQKVVEHILQ-HSGICINAGTTHHQ 1134
QY 733 CTPTWTGLYCTQRCPLGFYKIDCALICOQNGADC-DHIS-QCTCTGFMRGRCBK-- 788
DB 1135 CPLGYTSYCEEQ-----DECA-SNPCOHGATCNFISGIRCEVPGVGVGVEYVD 1187
QY 789 -----CPSGTYGCGRQICD-----CLNNGTC-DHNG-T 816
DB 1188 EQQNPQONGTCTIDLVNHRKCSPPRTGLCEBNIDECAGRHCLNGQCVDRIGYT 1247
QY 817 CYSPGKAGARCDQAGYIIVGNLN 840
DB 1248 CRCLPGEFGERCE-----GDIN 1264

```

RESULT 8
 NTCL_BRARE
 ID NTCL_BRARE STANDARD; PRT; 2437 AA.
 AC P46530;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 1 precursor.
 GN NOTCH1 OR NOTCH.
 OS Brachydanio rerio (zebrafish) (Danio rerio)..
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7935;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=94128602; PubMed=8297791;
 RA Bierkamp C., Campos-Ortega J.A.;
 RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
 its pattern of transcription during early embryogenesis.";
 RL Mech. Dev. 43:87-100(1993).
 CC - FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
 EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
 CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
 CC - SUBCELLULAR LOCATION: Type I membrane protein.
 CC - DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation
 CC stages. During gastrulation is differentially expressed,


```

FT DISULFID 627 636 BY SIMILARITY.
FT DISULFID 643 653 BY SIMILARITY.
FT DISULFID 648 662 BY SIMILARITY.
FT DISULFID 664 673 BY SIMILARITY.
FT DISULFID 680 691 BY SIMILARITY.
FT DISULFID 685 700 BY SIMILARITY.
FT DISULFID 702 711 BY SIMILARITY.
FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.

```

Query Match 14.6%; Score 987; DB 1; Length 2437;
 Best Local Similarity 24.8%; Pred. No. 2,1e-50;
 Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

```

QY 91 KSOCPPGESEGMCPKACDKVHRC-----IAINTQCEBNGMTNSSACD 140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 85 KDCVLAG--SDRLCLTPVNHACMNSPCRNQGTSLTLDTFTCCQPGMSGKTQLA-- 140

QY 141 GDHMGPHCTSRCCCKGALCNPTIG--ACHCAAGFRGRCEBQGTGNDCH--ORCQC 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 141 -----DPCASN--PCANGGQCAFESHYICTCPNPHGTCTQNDV-----NECAVSPSPC 188

QY 198 QNGATCDHVTGE--CRCPRTGTAFCEDL-----CPPGK 229
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 RNGGCIINEVSGSYLCRCPREYTGPHCQRLXQPLPSPCRSGGTQVSDTTHTCCLPGF 248

QY 230 HGPOCE-----QRC-----PCONG 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 TGQTCENHNVDCCTQHACENGSPCIDGINTYNCHCDKHWGTGYCTEDVDECELSPNACONG 308

QY 244 GVCHHVTG--ECSPSPGMMGTVCQ-----PCPEGRFGKN 276
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 GTCNHTIGFHCVCVNGMTGSDCSENIIDCASASCHGATCHDRVASFCECPHRTGLL 368

QY 277 C--SOEC--OCHNGGTCA--ATGQ--CHCSPGTYGERCO--DECPVGTGYVLCATGC 324
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 CHLDACISNCKOKGSMCDTNPVSGKALCTCPRGYTSACNODIDECISGAN----- 420

QY 325 QCVGKGCYHSGA--CLCEAGFAGEBCEARLCEBGLGICDKRCPCHELENTSHSHPMS 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 PCEHGGRCLENTKSGFCQCLQGYEBSPEMDV-----NKC-KSNPC--ONDAITCLDOI 470

QY 383 G--ECACRPGMSGLYCNET-----CSPGFGYACQ---QIC 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 471 GGFHCICMPGEGYVFCQINSDDCASQPLCNKCIDKINSFCEBPKGSGSLQYVDDEC 530

QY 414 S---CQNCADCDSVTK--CTCAGPFGKIDC-----STPCPL-----GTYGINSSR 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 531 ASTPCKNGAKCTDGPKNKTCCTPFGSGIHCELDINECASSPCHYGVCRDGVASFTCDCR 590

QY 456 CG-----CKNDVACSPVDS--CTCAGHGVDCSR----- 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 591 PGYGRLCETININECLSQPCRNHGTCTODRENAVYICTCPKGTGVNCEINIDCKRRKPCDY 650

```

```

QY 486 -----CPSGTWGFGCNLTQ-----CLNGACNLTIDG-----TCTCAPGWRG 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 651 GKCLDKINGIECVCEBPGISGSMCNINIDCALNFCNNGTC--IDGVNSFTCLCPDDEFRD 708

QY 523 EKC-----ELPCQDGTGYNL-----AERC--DCSHADGCHP 552
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 ATCLSQHNECSSNPCIHGSCIDQINSYRCVCEAGWGNCDININECLSNPCVNGTCKD 768

QY 553 -TTGH--CCLGLGMSGVHDSVCAEGRMP-----NCSL----- 583
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 MTSGLTLCRAGFGSPNCQNNINECASNPCLNQGSCIDVAGFKCNMCLPYTGEYCNVL 828

QY 584 -PCY--CKNGASCSPD-----GICE----- 601
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 APCSRPCKNGVCRESDPFSQNCNPAWGQTCEDVINECVARNPCTNGVCENLRGG 888

QY 602 ---CAPGFRCTTCOR-----ICSPGYGHRCSQTCPOCV 632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 889 FQCRNPFGTALENDIDCEPNPCNSGVCCODRVNGFVCLAGFRGERCAEDIDECV 948

QY 633 HSGPCHHITGLCCLPFTGALCNEVPSGRFGKNC--AGICT--CTNNGTCNPIDR 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 949 --SAPCRNGNCTDCVNSYT--CS--CPAGFSGINCEINPTDCTESSCFNGGTC--VDG 999

QY 687 ---SCOCYPMWISDC-----SQP-----CPPAHMGPNC---IH 714
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1000 ISSFCVCLPFTGNYCOHDVNECDSTRCCQNGSGSQDGTGYTKCPRHGTGLNCQSLVR 1059

QY 715 TCN--CHNGAFC--SAYDECKCTPGWTGLYCTQ----- 744
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1060 WCDSSPCNKGSSCWQOGASFQCCASGWTGYICDVPVSCEVAARQGVSAVILCRHAQ 1119

QY 745 -----RCPLGFGYKDC--ALICQ---CQNGACD--DHISG--OCTCRTFGRHRC 786
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1120 CYDAGNTHLCRQAGYTSYCOBOVDECPQPCQNGATCTYDGLGYSCEVPYGHMNC 1179

QY 787 QK-----CPSGTGFGYGRQICD-----CLNN 807
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1180 KEINECLSQPCQNGGTCTIDVNTYKSCPRGTGVCHEIDIDDSPSVDPLTGERCRNG 1239

QY 808 STC-DHITG--TCYCSPGKAKRCDQAGYIYGNLSLSTRSTALPADSY 854
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1240 GRCDVRVGGYGVCPAGFGERCE-----GDVNE--CLSDPCDPSGSY 1280

RESULT 9
NOTC_DROME STANDARD; PRT; 2703 AA.
AC P07207; P04154; O97458; Q9M4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus Notch protein precursor
GN N OR EG:140G11.1 OR EG:163A10.2 OR CG33936.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
RT product that shares homology with proteins containing EGF-like
RT repeats."
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
```


RT "Sequence of the notch locus of Drosophila melanogaster: relationship
RT of the encoded protein to mammalian clotting and growth factors.";
RT Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Agre A.R., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abiri J.F., Aghayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A.I., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertlora S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisli M., Kaulush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshine N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang M., Zhao Q., Zheng L.,
RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RT Science 287:2185-2195(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-Oregon-R;
RX MEDLINE-20196011; PubMed-10731137;
RA Benos P.V., Galt M.K., Ashburner M., Murphy L., Harris D.,
RA Bartell B.G., Ferraz C., Vidal S., Brun C., Demallies J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Motlier S., Galibert F., Bokkova D.,
RA Minna B., Kafatos F.C., Louis C., Siden-Klimos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modellell J., Peter A., Schoettler U., Jaekle H., Bucheton A.,
RA Belhert N., Dove G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitlou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valentini P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RT Science 287:2220-2222(2000).
RN [5]
RP SEQUENCE OF 2505-2611 FROM N.A.
RX MEDLINE-85099329; PubMed-2981631;
RA Wharton K.A., Yedvobnick B., Finnerty V.G., Artavanis-Tsakonas S.;
RT "Ops: a novel family of transcribed repeats shared by the Notch locus
RT and other developmentally regulated loci in D. melanogaster.";
RT Cell 40:55-62(1985).
RN [6]
RP SEQUENCE OF 1-8 FROM N.A.
RX MEDLINE-87257846; PubMed-3037327;
RA Kelley M.R., Kidd S., Berg R.L., Young M.W.;

RT "Restriction of P-element insertions at the Notch locus of Drosophila
RT melanogaster.";
RT Mol. Cell. Biol. 7:1545-1548(1987).
RN [7]
RP REVIEW.
RA Harris W.A.;
RT "Many cell types specified by Notch function.";
RT Curr. Biol. 1:120-122(1991).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands Delta
CC and Serrate to regulate cell-fate determination. Upon ligand
CC activation through the released notch intracellular domain (NICD)
CC it forms a transcriptional activator complex with Su(H)
CC (suppressor of hairless) and activates genes of the enhancer of
CC split locus. Essential for proper differentiation of ectoderm.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M16152; AAB59220.1; -
DR EMBL: M16153; AAB59220.1; JOINED.
DR EMBL: M16149; AAB59220.1; JOINED.
DR EMBL: M16150; AAB59220.1; JOINED.
DR EMBL: M16151; AAB59220.1; JOINED.
DR EMBL: K03508; AAA28725.1; -
DR EMBL: M13689; AAA28725.1; JOINED.
DR EMBL: K03507; AAA28725.1; JOINED.
DR EMBL: AE003426; AAF45848.2; -
DR EMBL: AL035436; CAB37610.1; -
DR EMBL: AL035395; CAB37610.1; JOINED.
DR EMBL: M12175; AAA74496.1; -
DR EMBL: M16025; AAA28726.1; -
DR PIR: A24420; A24420.
DR PIR: A24768; A24768.
DR PIR: A05267; A05267.
DR HSSP: P00740; 1EDM.
DR FLYBase: FBgn0004647; N.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001438; EGF-11.
DR InterPro: IPR002049; Lamln_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam: PR00008; EGF; 36.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBLDOD.
DR PRINTS: PR00011; EGF_LAMININ.
DR PRINTS: PR01452; NOTCH.
DR SMART: SM00248; ANK; 7.
DR SMART: SM00181; EGF; 36.
DR SMART: SM00179; EGF_CA; 35.
DR SMART: SM00004; NL; 2.
DR PROSITE: PS50088; ANK_REPEAT; 5.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 22.
DR PROSITE: PS00022; EGF_1; 34.
DR PROSITE: PS01187; EGF-2; 28.
DR PROSITE: PS01187; EGF_CA; 21.
KW Receptor; Transcription regulation; Activator; ANK Repeat; ANK
KW developmental protein; Neurogenesis; Receptor; ANK Repeat;
KW EGF-like domain; Transmembrane; Glycoprotein; Signal.

```

FT SIGNAL 1 44 POTENTIAL.
FT CHAIN 45 2703 NEUROGENIC LOCUS NOTCH PROTEIN.
FT DOMAIN 45 1745 EXTRACELLULAR (POTENTIAL).
FT TRANSHEM 1746 1766 POTENTIAL.
FT DOMAIN 1767 2703 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 58 95 EGF-LIKE 1.
FT DOMAIN 96 136 EGF-LIKE 2.
FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5.
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7.
FT DOMAIN 331 370 EGF-LIKE 8.
FT DOMAIN 372 408 EGF-LIKE 9.
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11.
FT DOMAIN 488 524 EGF-LIKE 12.
FT DOMAIN 526 562 EGF-LIKE 13.
FT DOMAIN 564 600 EGF-LIKE 14.
FT DOMAIN 602 637 EGF-LIKE 15.
FT DOMAIN 639 675 EGF-LIKE 16.
FT DOMAIN 677 713 EGF-LIKE 17.
FT DOMAIN 715 751 EGF-LIKE 18.
FT DOMAIN 753 789 EGF-LIKE 19.
FT DOMAIN 791 827 EGF-LIKE 20.
FT DOMAIN 829 865 EGF-LIKE 21.
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23.

```

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
 Best Local Similarity 26.8%; Pred. No. 7.3e-50;
 Matches 290; Conservative 102; Mismatches 287; Indels 395; Gaps 70;

```

OY 7 SCL-----SFILLCHMIGTASPLNLED--FNVCSEHSESYSTVQESTPHPPDQIYYTSC 60
DB 502 SCLDDPGTFRCVCMGFTGTQCEIDIDECQSNPC-----LNDQTC 541
OY 61 TDILNMFECTRHRSYRTAYRHEKTYRRKSQCCPFYESEMC-----VPHCADRCVHGR 117
DB 542 HDKINGFKS-----CALGF--TGARQINIDCQSPCKNR 576
OY 118 -----CIAPNTQCEPFGNGTNC--SACDGDHMGPHCTSRCKNCAKALCPITNG-ACH 168
DB 577 GICHDSIAGYSECPGPGTGTSCENINDCSN-----PCHRKCIDVNSFKCL 626
OY 169 CAGFRGM-----RCEDR-----CEGTYG-----NCHQRCQ 196
DB 627 CDPGTGYICQKQINECESNPOQFDGHCQDRGASYCCQAGTSGKNCENVNNECHSN-P 685
OY 197 CONGATC-DHYTG-ECRCPGTYGAFCEDLCPGKHGPOCEQRCPCONGYVC-HHYTG-E 252
DB 666 CNGGATCIDGINSYKCCQVPGFTGQHC-----KNVDECIS-SPCANNGVCIDOVNGYK 738
OY 253 CSCPFGMGTVC-----GQP-----GPEGRFGNCS-----OECO-- 282
DB 729 CECPRGFDAHCLSDVDEBCASNPVCNNEGRCEGNEFICHCPRGYTGARCELDIDECSSN 798
OY 293 -CHNGTC-DAATG-QCHCSFRTYGERCO--DECPVGTGYLVALETCQVNGKCY-HV 335
DB 799 PCQHGCTGYDKLMAFSCCCMGTYGQCKETNIDDC-----VTPNCGNGCTCIRY 848
OY 336 SG-ACLCBAGFAGECEARLCPREGLYIGK-DKRCPLHENTHSHCPMSG--EACKP 389
DB 849 NGKYCVCVPTFTGRDCESKMDP-----CASNRC-----KNEAKTPSSNFIIDESCCKL 897
OY 390 GWSGLYCHNETCSPFYGACQOICSCONGADCDVYTG--KCTCAFGFGIDC----- 439
DB 898 GYTRGYCDEDI-----DECSLSPRCGASCLANVPSYRCLCTGTGYSGRBCALINTDCA 951
OY 440 STPCP-----LGTYINC-----SSRC-----GCKNDVCSPPVDGS--CTCK 474
DB 952 SFPQNGGTCLDGIIDYDYSCLCVDFGRHRCETDINECLSPQONCATCSOYVNSTYCTCP 1011

```

```

OY 475 AGMHVDCSIR-----CPSGTWGFGCN-----LTQC-----CLN 503
DB 1012 LGFSINQNTDEDECTESSCLNGSGCITGINGYNCSLAGYSGANCQYKLLKCSNFCPLN 1071
OY 504 GGACNTLIDG--TCTCAPMGRGEC-----ELPCQDGTGYLNCARCDSCSHADGCHPT 553
DB 1072 GATCHEQNNEYTCHCPSGFTGKQCCSEYVDMCGQSPCENG-----ATCSQMK--HQF 1120
OY 554 TGHCRCLPGRNGVHD-----SYCABERNMPNCSLPFCYCKNGASCSP--DDGICECAPRG 608
DB 1121 S--CICSAGMTGKLCDDVOTISCODAADRKGLSLROLCKNNG--TKRDYGNSHVCYSQGYAG 1177
OY 609 TTCOR-----ICSPFGYGRHSQRCPOCV-----HSSGPCH 639
DB 1178 SYCQKEIDECQSPQONGTGRDLIGATECCQRRGFQONCELNIDDCAPRPNONGTICH 1237
OY 640 H--ITGLDCLPGFTGALC-----NEVCPSGRFRKNCAGICTTNTGCTPIDR---SCQC 690
DB 1238 DRYMNFSCSCPFTGIIICEINKDKCKPG-----ACHNNQSC--IDRVGGEFCVC 1285
OY 691 YPGWIGSNC-----SQPCP-----PAHMGPCITCN----- 717
DB 1286 QPEVGARCEEDINECLSNPSCNAGTLDVOLVNNHNCNRGRHGRHCEKRVDFCAQSP 1345
OY 718 CHNGAFCSAYDGECKCTPGMTGLYCTQRCPLGFYKDCALICQCONGACDHIS---GQC 774
DB 1346 CONGNCNIRQ-----SGHNCI--CNNGFYKNCCL-----SGDDCSNCRVNG 1389
OY 775 TCRTEFMGRHCEQKCPSTGYGCR---QICD---CLNSTCHITG--TCYSPGKNG 825
DB 1390 VVADEGFEYGRCE--CPRGTLGEHCHIDTIDECSPNFCQAGACEDLIDYELCLPSKWK 1447
OY 826 ARCD 829
DB 1448 KRCD 1451

```

RESULT 10
 NTG3_MOUSE
 ID NTG3_MOUSE STANDARD; PRT; 2318 AA.
 AC 061982;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR X Swiss Webster;
 RC MEDLINE=95001556; PubMed=7918097;
 RA Lardelli M., Dalstrand J., Lendahl U.;
 RT "The novel Notch homolog mouse Notch 3 lacks specific epidermal
 growth factor repeats and is expressed in proliferating
 neuroepithelium.";
 RT Mech. Dev. 46:123-136(1994).
 RN [2]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1664.
 RC MEDLINE=21523956; PubMed=11518718;
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 proteolysis.";
 RT J. Biol. Chem. 276:40268-40273(2001).
 RN [3]
 RP POST-TRANSLATIONAL PROCESSING.
 RC MEDLINE=21374376; PubMed=11459941;
 RA Mikutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 among mammalian Notch family members.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).


```

FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.
FT DISULFID 534 543 BY SIMILARITY.
FT DISULFID 550 560 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 592 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.

```

Query Match 14.58; Score 977.5; DB 1; Length 2318;
 Best Local Similarity 25.0%; Pred. No. 7.2e-50;
 Matches 321; Conservative 78; Mismatches 309; Indels 575; Gaps 69;

```

OY 14 LILCMITGATSPLENDNEDNCSHMSYSVTQESYHPHPDQIYYTSCDILNMFECTRHR 73
DB 30 LILLAGLGAAPCL-DSPCA-----NGRCHTHQ 59
OY 74 VSYRTAYHGEKTYRRKSQCCPGFYESGEMCVPHCADKCVHGRCAIAPNTCO----- 125
DB 60 PLEMA-----CLCLPGMV--GERC--QLEDPCHGSPAGAGVQSSVYAGTA 103
OY 126-----CEFGWGTCNSSACDGDHWPCHTSSRCCKNALCNPTG-----ACHCAAGFRG 175
DB 104 RFSRCRLGFGFGPDCSOP-----DPCVSR-PCVHGAPCS--VGPDRFACACPPGYOG 153
OY 176 WRCE-----DRCEGTGNDCHORCOCONGATCDHYTG--ECRCPPGYTAFCE--LPP 227
DB 154 OSCQSDIDECNSGT-----TCRHGCTCLNTPGSRRCQCPPLATGLLCEPNVYPCAP 204
OY 228 GKHGRCBORCPQONGVCH--VTGECSCPSGMMGTVCG--OPCEGFRGKNCQEC 281
DB 205 S-----PCRNNGTCROSSDVTYDCACLPFGEGNCEVNVDDCGHR----- 245
OY 222 OCHNGTC--DAATGQCNCSPYTERCO--DECPVGTGYLCAETCCVAGCKCYHS 336
DB 246 -CLNGTGVGVNTYNNCCQPPMTGQFCTEDVDECOL-----QPNACHNGGTCFNL 296
OY 337 G--ACICEAGFAGERCERL-----CPEGLYGKIC--DKRC 368
DB 297 GGHSCVCTNGWGTGESCSONIDDCATAVCFHATCHDRVASFACACPMGRTGLLCHLDAC 356
OY 369 ---PCHLENTSHCHPMGE--CACRPGMSGLYCNE----- 398
DB 357 VSNPCHEDALICTNPNVSRAICTCPFGTGACDQDVDECISGANPCEHLRCVNTQSGF 416
OY 399 -----TCSPPFYEGACQ-----QI 412
DB 417 LCQGRGYTGPRCETDVNECLSGPCRNQATCLDRIGQFTICJCMQAFETGYCEVDIDECOS 476
OY 413 CSCQNGADC-DSVTG-KCTCAGFKGIDC-----STPCPLGTG-----GNCSSNR 456
DB 477 SPCVNGVCKDRVNGFSCTCBGFGSGMKQLDVDECASTPCPNKGAKCYDQPGYTEC--RC 534
OY 457 G-----C-KNDVACSP-----VDG--SCTCRAGMHGVDS----- 483
DB 535 AEGFGTLCERNVDDCSPDCHHGRVCYGIASFSCACAPGTGICINSEQVDECRSQPCRY 594
OY 484 -----IRCPSTMGFCNLTG--QCLANG-----GACNTLIG-----TCTCAPGRNG 522
DB 595 GSKCLDLVKKYLCRCPPTGTGNCVENITDDCASNCPCTGVGR--DGINKRYDCVCPGFTG 652

```

```

OY 523 EXCEL-----PCODGTGTLNC-----AERDCSHADGCH 551
DB 653 PLCNWEINDECASSPCEGSGCYDGENGRHCLCPGSLPLCLPANHRAHRCPSHG-VCH 711
OY 552 PTTG--HCRCLPGMSGVHCDVYACABGRWPNCSLPCYCRNGASCSPDDGI--CECARGF 606
DB 712 DAPGFRVCERCEGSGRSCOSLA-----PDACESQPOAGGTCR-SBIGRFTCARGF 765
OY 607 KCTTCQRI-----CSPGYGHRCSQTCRCQCHSS-----G 636
DB 766 QGHQCEVLSPTCPSLCEHGHCESDPDRLLTVCSCPPMGWGRCCQDQVDECAASPCPHG 825
OY 637 PCNHTTG--LCDCLPGFALCNE-----YCPSGRFGKNC 669
DB 826 TCTNLPGNFRCLCHRGTYGPPCDQDIDCDPNPLCHGSSCQDGVGSFSCSLDGFAGPRC 885
OY 670 A-----GICP-----CTNNGTCNPIDR 686
DB 886 ARVDDECLSPGCPETCTDHYASFACAPPGYGFHCEIDLDPSCSPSCFNCGTC--VDG 943
OY 687 -----SCQCPGNIGSDC-----SQP-----CPRHMGPCNLIHTCN----- 717
DB 944 VSSFECILCRPGTYGTHCOYEADPCFSRPLHGLICNPYHPGFE--TCREGFTGSCQNP 1001
OY 718 -----CHNGAPCSAYDGECKCPGWTG-----LYCTQR----- 745
DB 1002 VDMCSQAPQONGRCVGTGATYICPPGWSGRGLCDIQSLPTEAANOMGVRLQCOEGCK 1061
OY 746 -----CPLEFYGRKC-----ALICQONGADCDHISG--QCTCRTGFMGRACE 786
DB 1062 CIDKRSYHCVPESRGTSHCHEYDPTAQCQHGRCRGVMGSGYCECPAGYAGDSCE 1121
OY 787 Q-----KCPSTGYGTC--ROICD-----CLNNS 808
DB 1122 DNIDECASQPCQNGSSCIDLVARYLCSPPGLTVLCEINEDDCDGLPSLDGVOCLHNG 1181
OY 809 TCDHITG--TCYCSPGMKARGCD 829
DB 1182 TCVDLVGFRNCPPGYTGELHCE 1204

```

RESULT 11
 NTC3_HUMAN STANDARD: PRT: 2321 AA.
 AC Q9UM47; Q9U618; Q9UPL3; Q9UEB3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97032728; PubMed=8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabriat H., Mouton P.,
 RA Alomovitch S., Domenga V., Cecillon M., Marechal E., MacLarek J.,
 RA Vayssiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissbach J.,
 RA Bach J.-F., Bousser M.-G., Tournier-Lasserre E.,
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 causing stroke and dementia".
 RL Nature 385:707-710(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Guzel M., Aravanis-Tsakonas S.,
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McGready P.M., Skowronski E., Adamson A.W.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
 RA Phan H., Velasco N., Garnes J., Danganan L., Poundstone P.,

CC
RT RT "Sequence analysis of an I.5 Mb olfactory receptor (OLFR) cluster in
RT 19p13.1.";
RL submitted (MAY-1998) to the EMBL/genbank/DBD databases.
RN [4]
RP VARIANTS CADASIL Y-49; C-71; C-90; C-110; C-133; C-141; R-146; C-153;
RP C-169; C-171; C-182; R-185; S-212; G-221; T-224; C-286; C-358;
RP C-578; C-128; C-985; C-1006; C-1031; C-1231 AND R-1261, AND VARIANTS
RP R-170; L-946; Q-1133; M-1183 AND A-2223.
RX MEDLINE=98048753; PubMed=9388399;
RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabriat H.,
RA Vayssiere C., Craud C., Maciazek J., Weissbach J., Bousser M.-G.,
RA Bach J.-F., Tourner-Lasserre E.;
RT "Strong clustering and stereocyped nature of Notch3 mutations in
RT CADASIL patients.";
RL Lancet 350:1511-1515(1997).
RN [5]
RX VARIANT CADASIL 114-GLY->PRO-120 DEL .
RX MEDLINE=20264473; PubMed=10802807;
RA Joutel A., Chabriat H., Vahedi K., Domenga V., Vayssiere C.,
RA Ruchoux M.M., Leys D., Bousser M.-G., Tourner-Lasserre E.;
RT "Splice site mutation causing a seven amino acid Notch3 in-frame
RT deletion in CADASIL.";
RL Neurology 54:1874-1875(2000).
RN [6]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
RA Banks A., Lemman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RL "Human ligands of the Notch receptor.";
AM Am. J. Pathol. 154:785-794(1999).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC jaggedl, jagged2 and deltal to regulate cell-fate determination,
CC upon ligand activation through the released notch intracellular
CC domain (NICD). It forms a transcriptional activator complex with
CC Rbf-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity).
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
CC tissues.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC the trans-Golgi network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- DISEASE: Defects in NOTCH3 are associated with CADASIL (Cerebral
CC autosomal dominant arteriopathy with subcortical infarcts and
CC leukoencephalopathy) which causes a type of stroke and dementia of
CC which key features include recurrent subcortical ischemic events
CC and vascular dementia.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LIN/NOTCH REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

FT DOMAIN 158 195 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 197 234 EGF-LIKE 5.
FT DOMAIN 236 272 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 274 312 EGF-LIKE 7.
FT DOMAIN 314 350 EGF-LIKE 8.
FT DOMAIN 351 389 EGF-LIKE 9.
FT DOMAIN 391 429 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 431 467 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 469 505 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 507 543 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 545 580 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 582 618 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 620 655 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 657 693 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 695 730 EGF-LIKE 18.
FT DOMAIN 734 770 EGF-LIKE 19.
FT DOMAIN 771 808 EGF-LIKE 20.
FT DOMAIN 810 847 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 849 885 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 887 922 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 924 960 EGF-LIKE 24.
FT DOMAIN 962 998 EGF-LIKE 25.
FT DOMAIN 1000 1034 EGF-LIKE 26.
FT DOMAIN 1047 1082 EGF-LIKE 27.

```

Query Match 14.4%; Score 974; DB 1; Length 2321.

Best Local Similarity 25.3%; Pred. No. 1.2e-49; Matches 304; Conservative 69; Mismatches 313; Indels 514; Gaps 69;

```

QY 107 PHCAD-KCVH-GRCI-APN--TCQCEPGNGTNCSSACGDDHMGPHCTSCCKNGAL 159
DB 41 PCLDGSFCANGRCGTOLPSREACICLPPGMVGERCOLE-DPCHSGP-CAGNGVQSSSV 98
QY 160 CNPTGACHCAAGFGMACE--DRCEQGTGNDCHQRCOCCNGATCDHYTGE----CRC 212
DB 99 AGTANFSRCRGRGFRGSPCSLPDCLSS--PCAHGARS--VGPDGRFLSC 146
QY 213 PPGYAGACE--DLCPGKHGPOCEORPCONGSVCHHTVG--DSCSPSGMGTVCQGP 267
DB 147 PPGYAGACRSRDVDECRAGE-----PCRHGSTCLNTPGSEFCQCPAGTGPLCENP 197
QY 268 CPEHFGKNCSEOCCHANGGTCDAA--TGQCHSPGYTGERCQ--DECP-----312
DB 198 AVP-----CAPS-PCRNGGTGRCGDLTYDOACLPFGEGNCEVNVDDCGHRLNGT 250
QY 313 ----VGTGYLC-----AETCQ-----CVNGKCYHVSQ--ACLEAGFAGER 349
DB 251 CVDGVTNVCQCPREMTGQFCTEDVDECOLOPNACHNGTGFNTLGGHSCVCVNGMTGES 310
QY 350 CEARL-----CPEGLYKIC--DKRC--PCHLEHTHSGP 380
DB 311 CSQNIIDCATVCFHGATCHDRVASFYCACPMGKTGLCHLDDACVSNPCHEDAICDTPNP 370
QY 381 MSGE--CAKRGWSGLYCN-----398
DB 371 VNGRAICTCPRGFTGACDDVDDECSIGANPCENHLGRCVNTGQSFLLCQGRGYTPRCET 430
QY 399 -----TCSPGFYGEACQ-----QICSCONGADC--DSVTG 426
DB 431 DVNECLSGPCRNOATCLDRIGQFTICIMAGFTGYCEVDIDECOSSPCVNGVCKDRYNG 490
QY 427 -KCTCAPGFKGIDC-----STPCPLGTY-----GINSRRG-----C-KNAV 463
DB 491 FSCICPSGFSSTGQLDVDECASTPCRNAGACVDPDGEYC--KCAESFEGTLCDRNVDD 548
QY 464 CSP-----VDG-----SCTCKAGWGVDS-----TRC 486
DB 549 CSPDCHGRGCVDGIAFSACAPGXTGTRESQVDECRSQPCRHRGKCLDLVYKLRG 608
QY 487 PSGTGFGPCNLTC-QCLNG-----GACNTLDG-----TTCAPAGWNGEKEL-----527
DB 609 PSGTGTVNCEVNIIDCASNPCTFGVCR--DGINRYDVCVOPGFTGPLCNVEINECCASPC 666

```

```

QY 528 ----PCODGTGYNL-----AERC--DCSHADGCHPTTG--HCRCILPGMS 564
DB 667 GEGGSCVDGNGEFRCLCPGSLPLCLPPSHPCAHEPSSH1-CYDAPGRVCCEPMS 725
QY 565 GVHCD-----VCAEGRMPNCSL--PCY----CKNGAS 592
DB 726 GPRCSQSLARACESPQRAGGTCSDDMGHCHTCPPVOGRQCELLSPCPNCEHGR 785
QY 593 CSPDGC--ICECAPGRFTTCQR-----ICSPGYGRH 623
DB 786 CESAPGQLPVCSCPOGMGPRCQDDVDECAAPACPGHICTNLAGSFSTCHGGYTPS 845
QY 624 CSQCPQC-----VSSGCHHTG--LCDCILPGFTGALCN-----658
DB 846 CDQDINDDPNCLNGSCQDGVGSFSCSLPGFAPGRCARDVDECLSNPCGPGTCDHV 905
QY 659 ----VCPSGFRGKNC-----AGICTNNGTNCNPIDR-----SCQCYPGMIGSPC-----699
DB 906 ASFTCTCPGXYGFRCEQDLDPCSPSSCFNGTGC--VDGVNSFSLCARGTGAHQCHEA 963
QY 700 ----SQP-----CPRAH-----WCPNC--IHTCN--CHNGAFCSAYDGECK 732
DB 964 DPCISRPCLHGVCSAAHPRGRCYCTLESFTGPOCTLVDMCSROPCCONGRCVOYGAYCL 1023
QY 733 CTGWTGLYCTQR-----CPLGYGKDC--755
DB 1024 CPGHSGRLCIRSLPCREAAQIGVRLQCLQAGGCVDEEDSSHYCVCPBGRGSHCEQ 1083
QY 756 ----ALICQNGADCDHISG--OCTCRTGFMGRHCEO-----787
DB 1084 EVDPLAPQCHGGTCRGYMGVMECLPGVNGDNCEDDVECASQPCQHGSSCIDLYAR 1143
QY 788 ----KCPSTYIGCRQICD-----CLNSTCDHITG--TYCSPGMKARCD 829
DB 1144 YLCSPPGTLGVLEINEDDCGPPPLDGSRCPLNCTVLYVGGFRCTCPPGYTLGRCE 1203

```

RESULT 12

NTC3_RAT

ID NTC3_RAT STANDARD; PRT: 2319 AA.

AC Q9R172;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).

GN NOTCH3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RA Haritunians T., Boulter J., Weimaster G., Schanen N.C.,

RT "Rattus norvegicus mRNA for Notch 3."

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP FUNCTION.

RX MEDLINE=21094508; PubMed=11182080;

RA Taniigaki K., Nogaki F., Takamashi J., Tashiro K., Kurooka H.,

RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent

RT neural progenitor cells to an astroglial fate."

RL Neuron 29:45-55(2001).

RN [3]

RP TISSUE SPECIFICITY.

RX MEDLINE=21331789; PubMed=11438922;

RA Iryin D.K., Zuercher S.D., Nguyen T., Weimaster G., Kornblum H.I.;

RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple

RT functional roles for the Notch-DSB signaling system during brain

RT development."

RL J. Comp. Neurol. 436:167-181(2001).

CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands

CC .Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.

CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC
 CC -1 SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC
 CC -1 TISSUE SPECIFICITY: Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within
 CC numerous cells throughout the CNS. It is more highly localized
 CC to ventricular germinal zones.
 CC
 CC -1 PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC
 CC -1 PTM: Phosphorylated (By similarity).
 CC
 CC -1 SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC
 CC -1 SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC
 CC -1 SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC
 CC -1 SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage and for commercial
 CC entitles requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sb.ch).
 CC -----
 CC EMBL: AF164486; A04653.2; -
 CC HSSP: P00740; 1EDM.
 CC
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR001881; EGF_Ca.
 CC InterPro: IPR001438; EGF_11.
 CC InterPro: IPR002049; Laminin_EGF.
 CC InterPro: IPR000800; Notch.
 CC Pfam: PF00008; EGF_34.
 CC Pfam: PF00023; ank; 6.
 CC Pfam: PF00066; notch; 3.
 CC PRINTS: PR00010; EGFBL00D.
 CC PRINTS: PR00011; EGF_LAMININ.
 CC PRINTS: PR01452; NOTCH.
 CC SMART: SM00248; ANK; 5.
 CC SMART: SM00179; EGF_Ca; 20.
 CC SMART: SM00001; EGF_like; 12.
 CC SMART: SM00004; NL; 3.
 CC PROSITE: PS50088; ANK_REPEAT; 4.
 CC PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 CC PROSITE: PS00010; ASX_HYDROXYL; 16.
 CC PROSITE: PS00022; EGF_1; 33.
 CC PROSITE: PS01186; EGF_2; 26.
 CC PROSITE: PS01187; EGF_Ca; 17.
 CC Receptor: Transcription regulation; Activator; Differentiation;
 CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
 CC SIGNAL 1 40
 CC CHAIN 41 2319 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.

FT CHAIN	1631	2319	NOTCH EXTRACELLULAR TRUNCATION (BY
FT CHAIN	1664	2319	SIMILARITY).
FT DOMAIN	41	1645	NOTCH INTRACELLULAR DOMAIN (BY
FT TRANSMEM	1646	1666	SIMILARITY).
FT DOMAIN	1667	2319	EXTRACELLULAR (POTENTIAL).
FT DOMAIN	41	79	POTENTIAL.
FT DOMAIN	80	120	CYTOPLASMIC (POTENTIAL).
FT DOMAIN	121	158	EGF-LIKE 1.
FT DOMAIN	160	197	EGF-LIKE 2.
FT DOMAIN	199	236	EGF-LIKE 3.
FT DOMAIN	238	274	EGF-LIKE 4.
FT DOMAIN	276	314	EGF-LIKE 5.
FT DOMAIN	316	352	EGF-LIKE 6.
FT DOMAIN	353	391	EGF-LIKE 7.
FT DOMAIN	393	431	EGF-LIKE 8.
FT DOMAIN	433	469	EGF-LIKE 9.
FT DOMAIN	471	507	EGF-LIKE 10.
FT DOMAIN	509	545	EGF-LIKE 11.
FT DOMAIN	547	582	EGF-LIKE 12.
FT DOMAIN	584	620	EGF-LIKE 13.
FT DOMAIN	622	657	EGF-LIKE 14.
FT DOMAIN	655	697	EGF-LIKE 15.
FT DOMAIN	697	732	EGF-LIKE 16.
FT DOMAIN	736	772	EGF-LIKE 17.
FT DOMAIN	773	810	EGF-LIKE 18.
FT DOMAIN	812	849	EGF-LIKE 19.
FT DOMAIN	851	887	EGF-LIKE 20.
FT DOMAIN	889	924	EGF-LIKE 21.
FT DOMAIN	926	962	EGF-LIKE 22.
FT DOMAIN	964	1000	EGF-LIKE 23.
FT DOMAIN	1002	1036	EGF-LIKE 24.
FT DOMAIN	1038	1084	EGF-LIKE 25.
FT DOMAIN	1086	1132	EGF-LIKE 26.
FT DOMAIN	1122	1160	EGF-LIKE 27.
FT DOMAIN	1162	1205	EGF-LIKE 28.
FT DOMAIN	1207	1246	EGF-LIKE 29.
FT DOMAIN	1248	1289	EGF-LIKE 30.
FT DOMAIN	1291	1337	EGF-LIKE 31.
FT DOMAIN	1337	1375	EGF-LIKE 32.
FT REPEAT	1384	1420	EGF-LIKE 33.
FT REPEAT	1425	1461	EGF-LIKE 34.
FT REPEAT	1469	1503	LIN/NOTCH 1.
FT REPEAT	1840	1869	LIN/NOTCH 2.
FT REPEAT	1873	1903	LIN/NOTCH 3.
FT REPEAT	1907	1936	ANK 1.
FT REPEAT	1940	1969	ANK 2.
FT REPEAT	1973	2002	ANK 3.
FT SITE	1573	1574	ANK 4.
FT DISULFID	44	56	ANK 5.
FT DISULFID	50	67	CLEAVAGE BY FURIN-LIKE PROTEASE (BY
FT DISULFID	69	78	SIMILARITY).
FT DISULFID	84	95	BY SIMILARITY.
FT DISULFID	89	108	BY SIMILARITY.
FT DISULFID	110	119	BY SIMILARITY.
FT DISULFID	125	136	BY SIMILARITY.
FT DISULFID	130	146	BY SIMILARITY.
FT DISULFID	148	157	BY SIMILARITY.
FT DISULFID	164	176	BY SIMILARITY.
FT DISULFID	170	185	BY SIMILARITY.
FT DISULFID	187	196	BY SIMILARITY.
FT DISULFID	203	214	BY SIMILARITY.
FT DISULFID	208	224	BY SIMILARITY.
FT DISULFID	226	235	BY SIMILARITY.
FT DISULFID	242	253	BY SIMILARITY.
FT DISULFID	247	262	BY SIMILARITY.
FT DISULFID	264	273	BY SIMILARITY.
FT DISULFID	280	293	BY SIMILARITY.
FT DISULFID	287	302	BY SIMILARITY.
FT DISULFID	304	313	BY SIMILARITY.
FT DISULFID	320	331	BY SIMILARITY.


```

FT DISULFID 325 340 BY SIMILARITY.
FT DISULFID 342 351 BY SIMILARITY.
FT DISULFID 357 368 BY SIMILARITY.
FT DISULFID 362 379 BY SIMILARITY.
FT DISULFID 381 390 BY SIMILARITY.
FT DISULFID 397 410 BY SIMILARITY.
FT DISULFID 404 419 BY SIMILARITY.
FT DISULFID 421 430 BY SIMILARITY.
FT DISULFID 437 448 BY SIMILARITY.
FT DISULFID 442 457 BY SIMILARITY.
FT DISULFID 459 468 BY SIMILARITY.
FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 480 495 BY SIMILARITY.
FT DISULFID 497 506 BY SIMILARITY.
FT DISULFID 513 524 BY SIMILARITY.
FT DISULFID 518 533 BY SIMILARITY.
FT DISULFID 535 544 BY SIMILARITY.
FT DISULFID 551 561 BY SIMILARITY.
FT DISULFID 556 570 BY SIMILARITY.
FT DISULFID 572 581 BY SIMILARITY.

```

Query Match 14.48; Score 969.5; DB 1; Length 2319;
 Best Local Similarity 25.18; Pred. No. 2.1e-49;
 Matches 303; Conservative 71; Mismatches 304; Indels 531; Gaps 62;

```

OY 107 PHCAD-----KCVHGRCAIPMTCCOEPGMMGNSSACDDGDMHGHCHTNRQCKMGA 158
D 42 PCCIDGSPCANGRCRTHQDPREACICLPGWVGEROLE-DPCHSGP-CAGRYVCSSV 99
OY 159 LCNPTGACHCAGFGRNCE-DRC-----BOGTVG-----188
D 100 VAGVARESCRLRGFRPDCSLPDPFCSSPCAHGAPSVSGDGRYACACPRGYGRNCRS 159
OY 189 --NCHQRCCQONATGHTG--ECRCPRGTGAFCD--LCPPGHNHGPCQECRCQ 241
D 160 DIDCRAGASCRCRGCTINTGSHFHCICLPGLTGLCNPTVPCAPS-----PCR 209
OY 242 NGVYCH--VTGECSPSGMGTVCG--OPCEGRGKNCSDCCCHNGSTC--DAAT 293
D 210 NGCTCRSSDVTYDCACLPGEQNCENVNDCGHR-----CLNGTGYDGYNT 259
OY 294 GQCHSPGYTERCQ--DECPVGTGYLCAETCCQVNGKCYHVS--ACLCGAGFAGE 348
D 260 YNCCCPREMTGOFCTEDVDECQL-----QPNACHNGGTCFNLGGHSCVCVNGMTGE 311
OY 349 RCEARL-----CPREGIYKIC--DKRC--PCHLENTSHCH 379
D 312 SCSONIDDCATAVCFHGATCHDRVASFYCACPMGKTGLCHLDACVSNPCHEDIAIDTN 371
OY 380 PMSGE--CAKPRGMSGLYCNB-----398
D 372 PVSRAITCPTCPFTGGACDDVDDECSIGANPCHEHLKRCVNTQGSFLCCGCGRTYGPKE 431
OY 399 -----TCSPGFYEGACQ-----QICSCONGADC-DSVT 425
D 432 TDVNECLSGPCRNATCLDIRIGOFICMAGFTGFCEVDIDEQSSPCVANGVCKDRVN 491
OY 426 G-KTCAAGFGIDC-----STPCPLGTY-----GINCSSRCG-----C-KDA 462
D 492 GFSCTCPBGSFSTCOLDVDECASTPCRNKAKVDPDGYEC--RCAAGFEFTLCERVD 549
OY 463 VCSF-----VDG-----SCTCKAGNHGVDCS-----IR 485
D 550 DCSHPCHHGRKVCVGLASFSCACAPGTGIGRCSQVDECRSPCRYGGKCLIDLVDKYICR 609
OY 486 CPSTGNGFGCNITC-QCLNG--GACNTLDG--TCTCAPRGKEXCEL-----527
D 610 CPPTTGVNCEVNIIDDAASNCTFGVCR--DGIRYDCVCPGTGTPLCANEIMECASSP 667
OY 528 -----PCODGTYGLNC-----AERDCSHADCSHPTTG--HCRCLPGW 563
D 668 CGEGSGCVDENGHFHCLCPGSLRPLCLPANHPCANRPSHSG-VCHDAPGGGFCVCDPGW 726

```

```

OY 564 SGVHDSVCARGRMGNPCSLPCYCKNGASCSPDDGI---CECAPGFRNCTGORI-----614
D 727 SGPRCSLSLA-----PDACESQPCOAGGCTCT-SBIGFHCICAPFGGHCQCEVLSPTCP 780
OY 615 -----CSRPFYGNHRSQTCPCQCVHS-----GPCNHTTG--LCDCL 648
D 781 LCEHGHCESDPDQUTVSCSPRMGMGPRCQDDVDECAACASPCHGTCTNLPSPRICH 840
OY 649 PFTGALCNE-----VCPSSGRFNKCA-----670
D 841 GGYTGPFCDODIDDCPNPCLNGSGCDQDVGFSFGSCSLSGFAGPRACARDVDECLSSPCP 900
OY 671 GICF-----CTNNGTCNPIDR-----SCOCYPMIG 696
D 901 GTCTIDHNASFTCTCPRGVGFHCEITDLDCSPSCFNNGTC--VDGVASFCLCPRTTG 958
OY 697 SDC-----SQP-----CPRAHMGNCIHTCN-----CHNGAF 723
D 959 THQYKVDPCFSRCLHGGICNPTHSGREC--TCREFTGNQCQNPVDMCSQAPCQNGR 1016
OY 724 CSAIDGECCTIPGWTG-----LYCIQR-----CPL 748
D 1017 CVOTGAYCICPEHWSGPLCDIPSLPCTEAAHMGVRLBOLQAGGQCIDKDSHYCPE 1076
OY 749 GFYKDC-----ALICCCONGADODHISG--OCTCRTFMGHRCEQ-----787
D 1077 GRMSHCQEVDPTAQPCQJGCTCRGMGVCCECPRTGYSDSCEDVDVDECAQPCONG 1136
OY 788 -----KCPSTGYGCRQICD-----CLNNTGCDHTG--TCYCS 820
D 1137 GSCIDLVAHYLCSPPGTLGVLICEINEDDCGPGSLDGLCLHNGTCTVDLVGFRNCNP 1196
OY 821 PGMKGARD 829
D 1197 PGYTGMRCE 1205

```

RESULT 13
 NTG4_HUMAN
 ID NTG4_HUMAN STANDARD; PRT: 2003 AA.
 AC Q99466; Q00306; Q99940; Q99458; Q9H358; Q9U119; Q9U10;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
 DE (Notch4).
 GN NOTCH4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND POLYMORPHISM OF POLY-LEU.
 RC TISSUE=Placenta;
 RX MEDLINE=97311416; PubMed=9168133;
 RA Suwaya K., Sasayama S.-I., Nobata J., Kimura T., Fukagawa T.,
 Nakamura Y., Ando A., Inoko H., Ikemura T., Mita K.;
 RT "Gene organization of human NOTCH4 and (CTG)n polymorphism in this
 human counterpart gene of mouse proto-oncogene Int3.";
 RL Gene 189:235-244(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
 RC TISSUE=Bone marrow, and Heart;
 RX MEDLINE=98360091; PubMed=9693032;
 RA Li L., Huang G.M., Banta A.B., Deng Y., Smith T., Dong P.,
 Friedman C., Chen L., Trask B.J., Spies T., Rowen L., Hood L.;
 RT "Cloning, characterization, and the complete 56.8-kilobase DNA
 sequence of the human NOTCH4 gene.";
 RL Genomics 51:45-58(1998).
 RN [3]
 RP SEQUENCE OF 1-503 FROM N.A., AND VARIANTS GLN-117 AND GLN-317.
 RA Miyagawa T., Tokunaga K., Hojho H.;
 RT "Human notch4 gene variant.";


```

FT DISULFID 74 103 BY SIMILARITY.
FT DISULFID 105 114 BY SIMILARITY.
FT DISULFID 122 133 BY SIMILARITY.
FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.
FT DISULFID 205 220 BY SIMILARITY.
FT DISULFID 222 231 BY SIMILARITY.
FT DISULFID 238 249 BY SIMILARITY.
FT DISULFID 243 262 BY SIMILARITY.
FT DISULFID 264 273 BY SIMILARITY.
FT DISULFID 280 291 BY SIMILARITY.
FT DISULFID 285 300 BY SIMILARITY.
FT DISULFID 302 311 BY SIMILARITY.
FT DISULFID 318 332 BY SIMILARITY.
FT DISULFID 326 341 BY SIMILARITY.
FT DISULFID 343 352 BY SIMILARITY.

```

Query Match 14.2%; Score 959.5; DR 1; Length 2003;

Best Local Similarity 26.5%; Pred. No. 7.1e-49; Indels 437; Gaps 64;

Matches 298; Conservative 64; Mismatches 324;

```

QY 94 CCEGFEYSGEMCYPHCADKC-----VHGRCIAPNT-----CCEPFGMGTCNSACDGDH 143
DB 105 CLRGF--TGERCQAKLEDPERPSCSKRGCHIOASGRPCSCMFGWTGEGQOLR----- 157
QY 144 WPHCTSRCCCKNGALCNPTG--ACHCAAGFRGWCE--DRCEOG---TYGNDCHQ- 193
DB 158 --DFCSAN--PCVNGVCLATYPTQIOCHCPPEFGHACERDVNCFODPPCPKGTSCNHT 214
QY 194 -----RCQ-----CONGATC-----DHVTCRCRPPRYTAFCF 222
DB 215 LGSFQCLPYVGEGPCELRAGCPRPGRCSNGGCOLMPEKDTFHLCLCPPEGTGDC 274
QY 223 -----DLCPPGRKH-----POCEORCP--CONGVCHH 248
DB 275 VNPDNVSHCCQNGTCQDGLDITTCCLPFTWTGMDCSEVDCECTGPPHCRNGGTCQN 334
QY 249 VTG--ECCSPSGMWTVCGP-----CPEGRFGKNCSE- 280
DB 335 SAGSFHCVCYSGWGTGSCENLDICAIATCAPGSTCIDRVGSFCLCPERTGLCHLED 394
QY 281 -C---QCHNGTC--DAATGQ--CHCSPGTGERCQ---DECPRYGTGYLCAETCCQVNG 329
DB 395 MCLSPCHGDAOCSTNLTSTLCQPGYSGPTCHQDDECLMAOQ-----PSPCEHG 449
QY 330 GKCYHVSQA--CLCEAGFAGEGCEAR-----LCPEGLYGIK 363
DB 450 GSCLTNTPGSFNCLCPRYTYSRCEADHNECLSPCRHSGSTCLDLATFHCLCPPEGLBQL 509
QY 364 CD---KRC---PCHLENTSHCHPMG--ECACKRPMGSLYCN----- 398
DB 510 CEVETNECASAPC--LHNHADCHDLNFGFCILPGFSGTCEBDEIDECSSPCANGGCOQ 567
QY 399 -----TCSRGFTGACQ--QIC---SCQNGADCDSTGK--CTCAGGFKIDCSTP 442
DB 568 DQPGAFHCKCLPGEFGRCTEVEDECLSDPCPVGASCLDIPGAFELCPGSGFTGOLCEVP 627
QY 443 -----CPLGTGT-----NCSRCG--CKNDVACSPYDSCGTC 473
DB 628 LCAPNLCQPKQICDKDKANDKANCCLPDCSPGACAPEDNCTCHHGHQCR-----SSCYC 679
QY 474 KAGHGVDC-----SIRCPSGTWGFGCN--LTCQ---CL 502
DB 680 DVGWTGPECEAELEGACISAPCAGGTCTYPPPSGYNCTCPRTGYPTCSSEMTACHSPCL 739
QY 503 NGGACNTLUG--TCTCAPGRGENCE--LPC-----QDGTGLNCA----- 539
DB 740 NGGSCNPSPGYCTCPBSHTGPOCQSTNDYCVSAPCFNGGTCVNRGTPSCCLAMGEOG 799

```

```

QY 540 ERCD-----CSHADGCH--PTTGHCRLCLPGMSGVCHDS----- 570
DB 800 PROEGKLRPCSCADSPCNRNRTCCODSPGPRCLCTGTGTGSSCCQTLMDLCAQKRCPRNSHC 859
QY 571 -----VCABGRMGPNCSLP-----CYCKNGASCSPDDG---ICEC 602
DB 860 LQTPSPFHCLCQGWGTGPLCNLPLSSCQKAALSGQIDVSSLCHNGGLC--VDGSPSYFCHC 918
QY 603 APGFRGTQCR-----ICSPGFYGHRSQRCPCQVHSSGP 637
DB 919 PPGFQSLCODHVNPCESRPCQNGATCMAPSGYLQCAAGYDQNCSEKLDAC--QSP 976
QY 638 CHHITGCLDCLPFTGALCNEVCPSGRFGKNAC-----ICTCTNNGTCNPIDRS-- 687
DB 977 CHN--HGCTPKRG--GPHC--ACPPRGVGLRCGDDVECLDQCHPTGTAACHSLANAFY 1031
QY 688 CQCPCWITGSDCSQPCPRAHMGPNCHTCCNHGAFCSAYDG---CKCTPGWTGLYC 742
DB 1032 COCLPHTGQWCEVEIDPCHSOP-----CFHGGTCETATAGSPLGFIHCPRGFEGPTC 1084
QY 743 TORCP--LGFYGNKCALICQONGADC-----DHISGCTCCTGFMGRHC--EQKCPSGTYG 795
DB 1085 SHRAPSCGFH-----HCHHGGLCLPSRPGPPPCACLSYGIGPDCULPPAPK---- 1132
QY 796 YGCRQIDCLNNTSCDHTG-----TCYCSPGMKARGCQDAG 832
DB 1133 -GGGPPSPCLYNGSCSETTGLGGRGFRGSCPHSSPGRCKRG 1174

```

```

RESULT 14
FBP1_STRPU STANDARD; PRT; 1064 AA.
ID FBP1_STRPU
AC P10079;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibropellin I precursor (Epidermal growth factor-related protein 1)
DE (UEGF-1).
GN BGF1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_Taxid-7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-90112459; PubMed-2514273;
RA delgadillo-Reynoso M.G., Rollo D.R., Hursh D.A., Rafi R.A.;
RT "Structural analysis of the uegf gene in the sea urchin
RT strongylocentrotus purpuratus reveals more similarity to vertebrate
RT than to invertebrate genes with EGF-like repeats."
RL J. Mol. Evol. 29:314-327(1989).
RN [2]
RP SEQUENCE OF 279-476 AND 781-1064 FROM N.A.
RX MEDLINE-87319677; PubMed-3498216;
RA Hursh D.A., Andrews M.E., Rafi R.A.;
RT "A sea urchin gene encodes a polypeptide homologous to epidermal
RT growth factor."
RL Science 237:1487-1490(1987).
RN [3]
RP AVIDIN-LIKE DOMAIN.
RX MEDLINE-89196806; PubMed-2784773;
RA Hunt L.T., Barker W.C.;
RT "Avidin-like domain in an epidermal growth factor homolog from a sea
RT urchin."
RL FASEB J. 3:1760-1764(1989).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE-91285254; PubMed-2060714;
RA Bisgrove B.W., Andrews M.E., Rafi R.A.;
RT "Fibropellins, products of an EGF repeat-containing gene, form a
RT unique extracellular matrix structure that surrounds the sea urchin
RT embryo."

```

RL Dev. Biol. 146:89-99(1991).
 CC -1- FUNCTION: FORM THE APICAL LAMINA, A COMPONENT OF THE EXTRACELLULAR
 CC MATRIX.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR, IN VESICLES IN THE CYTOPLASM
 CC OF UNFERTILIZED EGGS, THEN TO THE BASE OF THE HYALIN LAYER
 CC THROUGHOUT DEVELOPMENT AND FINALLY IN THE APICAL LAMINA IN LATE
 CC EMBRYOS AND EARLY LARVAE.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1A (SHOWN HERE) AND 1B: ARE
 CC PRODUCED BY ALTERNATIVE SPLICING. THE SMALL FORM (1B) LACKS 8 EGF
 CC REPEATS.
 CC -1- DEVELOPMENTAL STAGE: MODERATE LEVELS IN UNFERTILIZED EGGS AND
 CC DRINK EARLY CLEAVAGE, THEN RAPIDLY INCREASES IN ABUNDANCE BETWEEN
 CC LATE MORULA AND MESENCHYME BLASTULA STAGES TO MAXIMAL LEVELS
 CC MAINTAINED THROUGH SUBSEQUENT STAGES. EXPRESSED BOTH MATERNALLY
 CC AND ZYGOTICALLY.
 CC -1- SIMILARITY: CONTAINS 21 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 CUB DOMAIN.
 CC -1- SIMILARITY: THE C-TERMINAL DOMAIN OF THIS PROTEIN IS SIMILAR
 CC TO AVIDIN/STREPTAVIDIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L08692; AAA62164.1; -
 DR EMBL; L08692; AAA62163.1; -
 DR EMBL; X17530; CA35571.1; -
 DR EMBL; M17421; AAA30050.1; -
 DR EMBL; X17533; CA35573.1; -
 DR PIR; A29316; A29316.
 DR HSSP; P01132; 1EGF.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000086; Avidin.
 DR InterPro; IPR000859; CUB_domain.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF-IT.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 21.
 DR Pfam; PF00431; CUB; 1.
 DR Pfam; PF01382; Avidin; 1.
 DR PRINTS; PR00010; EGFBLD.
 DR PRINTS; PR00011; EGFBLD.
 DR SMART; SM00042; CUB; 1.
 DR SMART; SM00179; EGF_CA; 18.
 DR PROSITE; PS00001; EGF-like; 3.
 DR PROSITE; PS00010; ASX_HYDROXYL; 19.
 DR PROSITE; PS00022; EGF_1; 19.
 DR PROSITE; PS00577; AVIDIN; 1.
 DR PROSITE; PS01180; CUB; 1.
 DR PROSITE; PS01186; EGF-2; 19.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Biotin; Alternative splicing; EGF-like domain; Repeat; Signal;
 KW Glycoprotein; Calcium-binding.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 1064 FIBROPELLIN I.
 FT DOMAIN 20 55 EGF-LIKE 1.
 FT DOMAIN 62 175 CUB.
 FT DOMAIN 176 212 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 214 250 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 252 288 EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 290 326 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 328 364 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 366 402 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 404 440 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 442 478 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 480 516 EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 518 554 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).

FT DOMAIN 556 592 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 594 630 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 632 668 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 670 706 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 708 744 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 746 782 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 784 820 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 822 858 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 860 896 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 898 934 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 936 1064 AVIDIN-LIKE.
 FT DISULFD 23 34 BY SIMILARITY.
 FT DISULFD 28 43 BY SIMILARITY.
 FT DISULFD 45 54 BY SIMILARITY.
 FT DISULFD 180 191 BY SIMILARITY.
 FT DISULFD 185 200 BY SIMILARITY.
 FT DISULFD 202 211 BY SIMILARITY.
 FT DISULFD 218 229 BY SIMILARITY.
 FT DISULFD 223 238 BY SIMILARITY.
 FT DISULFD 240 249 BY SIMILARITY.
 FT DISULFD 256 267 BY SIMILARITY.
 FT DISULFD 261 276 BY SIMILARITY.
 FT DISULFD 278 287 BY SIMILARITY.
 FT DISULFD 294 305 BY SIMILARITY.
 FT DISULFD 299 314 BY SIMILARITY.
 FT DISULFD 316 325 BY SIMILARITY.
 FT DISULFD 332 343 BY SIMILARITY.
 FT DISULFD 337 352 BY SIMILARITY.
 FT DISULFD 354 363 BY SIMILARITY.
 FT DISULFD 370 381 BY SIMILARITY.
 FT DISULFD 375 390 BY SIMILARITY.
 FT DISULFD 392 401 BY SIMILARITY.
 FT DISULFD 408 419 BY SIMILARITY.
 FT DISULFD 413 428 BY SIMILARITY.
 FT DISULFD 430 439 BY SIMILARITY.
 FT DISULFD 446 457 BY SIMILARITY.
 FT DISULFD 451 466 BY SIMILARITY.
 FT DISULFD 468 477 BY SIMILARITY.
 FT DISULFD 484 495 BY SIMILARITY.
 FT DISULFD 489 504 BY SIMILARITY.
 FT DISULFD 506 515 BY SIMILARITY.
 FT DISULFD 522 533 BY SIMILARITY.
 FT DISULFD 527 542 BY SIMILARITY.
 FT DISULFD 544 553 BY SIMILARITY.
 FT DISULFD 560 571 BY SIMILARITY.
 FT DISULFD 565 580 BY SIMILARITY.
 FT DISULFD 582 591 BY SIMILARITY.
 FT DISULFD 598 609 BY SIMILARITY.
 FT DISULFD 603 618 BY SIMILARITY.
 FT DISULFD 620 629 BY SIMILARITY.
 FT DISULFD 636 647 BY SIMILARITY.
 FT DISULFD 641 656 BY SIMILARITY.
 FT DISULFD 658 667 BY SIMILARITY.
 FT DISULFD 674 685 BY SIMILARITY.
 FT DISULFD 679 694 BY SIMILARITY.
 FT DISULFD 696 705 BY SIMILARITY.
 FT DISULFD 712 723 BY SIMILARITY.
 FT DISULFD 717 732 BY SIMILARITY.
 FT DISULFD 734 743 BY SIMILARITY.
 FT DISULFD 743 761 BY SIMILARITY.
 FT DISULFD 750 761 BY SIMILARITY.
 FT DISULFD 755 770 BY SIMILARITY.
 FT DISULFD 772 781 BY SIMILARITY.
 FT DISULFD 788 799 BY SIMILARITY.
 FT DISULFD 793 808 BY SIMILARITY.
 FT DISULFD 810 819 BY SIMILARITY.
 FT DISULFD 826 837 BY SIMILARITY.
 FT DISULFD 831 846 BY SIMILARITY.
 FT DISULFD 848 857 BY SIMILARITY.
 FT DISULFD 864 875 BY SIMILARITY.
 FT DISULFD 869 884 BY SIMILARITY.
 FT DISULFD 886 895 BY SIMILARITY.
 FT DISULFD 902 913 BY SIMILARITY.
 FT DISULFD 907 922 BY SIMILARITY.

Query Match 14.2% Score 954.5; DB 1; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 7.6e-49;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

FT DISULFD 924 933 BY SIMILARITY.
 FT CARBOHYD 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 136 136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 851 851 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VASPLIC 477 780 MISSING (IN ISOFORM IB).
 FT CONFLICT 279 279 L -> S (IN REF. 2).
 SO SEQUENCE 1064 AA; 112072 MW; 2E569CA012ED6D09 CRC64;

30 DPNVCSHWSYVTVQESYPRHFDQIYYTSCDILNMFKCTHRVSYRTAYRHGKMTMR 89
 181 DPNLCQNG-----ACCTDLYNDACT----- 201
 90 RKSQCCPEFYSEGENC---VPHCA-DKCVHGRCLAPN---TCQEPGNGTNCSSACDG 141
 202 ---CPGPF--TGRNCEIDIDEACADPCONGACVDYGVYCNCGVPCFDECENNIN- 254
 142 DHMHPCHSRCKNGALCNPTGA---CHCAAGFRGMRCE---DREGGTGNDCHOR 194
 255 -----ECAS-SPCLNGGIC--VDGVNMEFCTCLAGFTVRCVENIDECAS----- 296
 195 GOCONGATC-DHYTG-ECRCPGYTGATCEDLCPRGKHGPOCEORCPONGGVCHHTYGE 252
 297 APCONGGICIDGINGTYTSCPLGFSGDNCE-----NDDECS--IFCLNGTCTVDLYNA 349
 233 ---CSCPSSMGTVCQPCPEGRFGNGSQECCNGHGTG-DAATG-QCHSPGYTGERCO 308
 350 YMCVAPMTGFTCADNIDE-----CA-SAPCONGVCLDGVNMYMDCQGYGTGHE 402
 309 ---DECPVGTGVLCAETNCCQVNGGKCYH-VSG-ACICEAGRGRCARLCBELYGRK 363
 403 TDIDEC-----APPCONGGDCVDGVNMYCTICAPGFDGNCE----- 440
 364 CDKRCPCHELENTHSC--HPMSGECAKPGWSGLYCNETCSPEFYGEACQ---QICS---C 415
 441 -----NNIDECASRPCONGAYCVDGVNGFVC--TCSAGYTVLCTDINECASMPIC 489
 416 ONGADC-DSYTGK-CTCAPRGIDCSTPCPLGTYGINSRCKGCKNDAYCS-PYDG-SC 471
 490 LMGVCTDLVNGYICTCAAGEGTGCTD-----ECAS-FPCONATCTDQVNGYVC 542
 472 TCKAGMAYDVC-----SIRCPSG-----TWGFGCNL----- 500
 543 TCVPEYTVLCTDINECASFPCLNGGTCNDQVNGYVCACADIVSTCTEDRDECASAP 602
 501 CLNGGAC-NTLDG-TCTCAPRGKCEKCEL-----PCQDGTGILNCAERDCSHADGC 550
 603 CLNGGACADVNGFYCTCLPMEGTNCEINDECASSPCOMNG--GL-CYDQVN-SYV--- 655
 551 HPTTHCHCLPQMSGVHCDSCVACBRKWPNCSLPCYCKNGASC--SPDGCICECAPRG 608
 656 -----CCLPQGTGTHGTEIDECASAP-----CLNGGQCIDRVDSYECVCAGYTA 702
 609 TTCC-----RSCPSGYHRSQTCPOCVHSGPC----- 638
 703 VRCQINIDECASAPONGGVYGVNGVYCNAPGTYDNCETIDEDEC--ASMPCLNGA 760
 639 ---HHITG-LCDCLPGFTGALC-----NEVCPSGRFGKNCAGICTCTNNGTCPIDRS 687
 761 CLEWNGYTCQCVAGYTVICETDIDECASAPCONG-----GYCTIDININ- 807
 668 CQCYGMWIGSDSQPCPAHMGPNCTHPCNHNAGFASAYDG-----ECKCPGWTGLYCT 743
 808 CACVGFSGNSCETNIDECASDP-----CLNGGIC--VQGVNFGVQCQPPNYSGYTCE 858
 744 QRCPLGFGKCALICOCONGADCDHISGO--CTCRGTGFMGNHC---QKCSGYTGGC 798
 859 -----ISLDACRMPCONGATCVNAGADYCECVPGAGNCHIDINECAS----- 904
 799 ROICDCLNNSTC-DHITG-TCYCSPGMKGARCDQAGVLIYVNL-----NSLSRTSTA 848

DB 905 ---LPCQNGGCLCIDIDAGYTCOCRLGYIGVNCCEEVGFCDLEGMNTECNDQYTIKRTSTG 961
 QY LPADSYQIAGIAGIITLVVLFLLALPIYRHKKRGESSPANTYTPARAYVADTTI 908
 DB 962 M-----MIGDMYNERALGYAAPVYVGYASN-----NYDPPS 995
 QY 909 SG-TLPHSNGGNANS 922
 DB 996 FGVTVVRDNGOSTTS 1010

RESULT 15
 NTCA_MOUSE STANDARD: PRT: 1964 AA
 AC P31695; 062389; 062390; 035442; 09RLW9; 088314; 088316; 09RLX0;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 4 precursor (Notch 4).
 OS [contamin: Transforming protein Int-3].
 GN NOTCH4 OR INT3 OR INT-3.
 OC Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92194507; PubMed=1312643;
 RX Robbins J., Blondel B.J., Gallahan D., Callahan R.;
 RT "Mouse mammary tumor gene Int-3: a member of the notch gene family
 transforms mammary epithelial cells."; J Virol. 66:2594-2599(1992).
 RL [2]
 RP REVISIONS, SEQUENCE FROM N.A.
 RA MEDLINE=97294599; PubMed=9150355;
 RX Gallahan D., Callahan R.;
 RT "The mouse mammary tumor associated gene INT3 is a unique member of
 the NOTCH gene family (NOTCH4)."; Oncogene 14:1883-1890(1997).
 RL [3]
 RP SEQUENCE FROM N.A.
 RA TISSUE-Lung, and Testis;
 RX MEDLINE=96281668; PubMed=6681805;
 RA Uytendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
 RT "Notch4/Int-3, a mammary proto-oncogene, is an endothelial
 cell-specific mammalian Notch gene."; Development 122:2251-2259(1996).
 RL [4]
 RP SEQUENCE FROM N.A.
 RA Rowen L., Mahairas G., Qin S., Ahern M.E., Dankers C., Lasky S.,
 RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
 RT "Sequence of the mouse major histocompatibility locus class III
 region."; J Biol Chem 269:16111-16116(1994).
 RL [5]
 RP Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 1436-1600 FROM N.A.
 RX MEDLINE=99252212; PubMed=10233982;
 RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
 RT "Intracisternal type A particle-mediated activation of the Notch4/Int3
 gene in a mouse mammary tumor: generation of truncated Notch4/Int3
 mRNAs by retroviral splicing events."; J Virol. 73:5166-5171(1999).
 RN [7]
 RP FUNCTION.
 RX MEDLINE=21244657; PubMed=11344305;
 RA Uytendaele H., Ho J., Rossant J., Kitajewski J.;
 RT "Vascular patterning defects associated with expression of activated
 Notch4 in embryonic endothelium."; Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
 RN [7]
 RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
 OF VAL-1463.

RA MEDLINE-21523956; Pubmed-11518718;
 RA Sakena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Notch homologs (N-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273 (2001).
 RN [8]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE-21374376; Pubmed-11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity). May regulate branching
 CC morphogenesis in the developing vascular system.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(E) which are probably linked by disulfide
 CC bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
 CC kidney, and at lower levels in the ovary and skeletal muscle. A
 CC very low expression is seen in the brain, intestine, liver and
 CC testis.
 CC -1- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
 CC embryonic development from 9.0 d.p.c.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(E). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- DISEASE: Loss of the extracellular domain causes constitutive
 CC activation of the Notch protein, which leads to hyperproliferation
 CC of granular epithelial tissues and development of mammary
 CC carcinomas.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 29 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M60456; AAB38377.1; -;
 DR EMBL: UA3691; AAC52630.1; -;
 DR EMBL: UA3691; AAC52631.1; -;
 DR EMBL: AF030001; AAB82004.1; -;
 DR EMBL: AF030001; AAB82004.1; -;
 DR EMBL: AB016771; BAA32281.1; ALT_SEQ.
 DR EMBL: AB016772; BAA32283.1; ALT_INIT.
 DR EMBL: AB016773; BAA32284.1; ALT_INIT.
 DR EMBL: AB016774; BAA32285.1; -;
 DR PIR: A38072; TVMT3.
 DR HSP: P08709; 1BF9.
 DR MGI: 107471; Notch4.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.

DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PF00008; EGF_27.
 DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 2.
 DR PRINTS: PRO1415; ANKYRIN.
 DR PRINTS: PRO0010; EGFLOOD.
 DR PRINTS: PRO1452; NOTCH.
 DR SMART: SM00248; ANK; 5.
 DR SMART: SM00179; EGF_CA; 11.
 DR SMART: SM00001; EGF-like; 15.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 11.
 DR PROSITE: PS00022; EGF_1; 28.
 DR PROSITE: PS0186; EGF_2; 21.
 DR PROSITE: PS0187; EGF_CA; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
 KW SIGNAL
 FT CHAIN 1 1964
 FT CHAIN 21 1964
 FT CHAIN 1411 1964
 FT CHAIN 1428 1964
 FT CHAIN 1463 1964
 FT DOMAIN 21 1443
 FT TRANSLEM 1444 1464
 FT DOMAIN 1465 1964
 FT DOMAIN 21 60
 FT DOMAIN 115 112
 FT DOMAIN 153 185
 FT DOMAIN 191 229
 FT DOMAIN 231 271
 FT DOMAIN 273 309
 FT DOMAIN 311 350
 FT DOMAIN 352 388
 FT DOMAIN 389 427
 FT DOMAIN 429 470
 FT DOMAIN 472 508
 FT DOMAIN 510 546
 FT DOMAIN 548 584
 FT DOMAIN 586 622
 FT DOMAIN 623 656
 FT DOMAIN 658 686
 FT DOMAIN 688 724
 FT DOMAIN 726 762
 FT DOMAIN 764 800
 FT DOMAIN 803 839
 FT DOMAIN 841 877
 FT DOMAIN 878 924
 FT DOMAIN 926 964
 FT DOMAIN 964 1000
 FT DOMAIN 1002 1040
 FT DOMAIN 1042 1081
 FT DOMAIN 1083 1122
 FT DOMAIN 1126 1167
 FT DOMAIN 1168 1208
 FT REPEAT 1209 1242
 FT REPEAT 1243 1282
 FT REPEAT 1283 1657
 FT REPEAT 1658 1691
 FT REPEAT 1692 1724
 FT REPEAT 1725 1757
 FT ANK 1.
 FT ANK 2.
 FT ANK 3.
 FT ANK 4.

Query Match 14.1%; Score 951.5; DB 1; Length 1964;
 Best local Similarity 26.5%; Pred. No. 2e-48;
 Matches 298; Conservative 67; Mismatches 316; Indels 445; Gaps 65;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:47:55 ; Search time 33.0243 Seconds

(without alignments)
3318.560 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744

Sequence: 1 MVLINSLSCSLFILLCHMT.....SSPKQEDSGSSSSSSSSSR 1140

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: PIR-73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1958	29.0	1574	2 T13954	MEGF6 protein - ra
2	1900	28.2	1620	2 T27283	hypothetical prote
3	1805.5	26.8	1111	2 T26972	hypothetical prote
4	1036	15.4	2524	2 A35844	notch protein - Af
5	1032	15.3	2555	2 A40043	notch protein homo
6	1028	15.2	2531	2 A46019	Notch-1 protein -
7	1024	15.2	2531	2 A18188	notch protein homo
8	996	14.8	2471	2 A49128	cell-fate determin
9	987	14.6	2437	2 S42612	transmembrane prot
10	978.5	14.5	2703	1 A24420	notch protein - fr
11	977.5	14.5	2318	2 S45306	notch 3 protein -
12	974	14.4	2321	2 S78549	notch3 protein - h
13	964.5	14.3	2531	2 T31070	notch homolog - se
14	954.5	14.2	1064	2 A40136	fibropellin Ia - s
15	952.5	14.1	1964	2 T09059	notch4 - mouse
16	939.5	13.9	1203	2 A49175	Notch B protein -
17	920.5	13.6	4006	2 T09070	probable tenascin
18	915.5	13.6	2352	2 T30201	Notch homolog prot
19	914.5	13.6	3566	1 A40701	tenascin-X precurs
20	881.5	13.1	4135	2 T42629	tenascin-X - bovin
21	775.5	11.5	2139	2 A35672	tenascin-C - human
22	769	11.4	2201	2 A32160	tenascin-C - human
23	767.5	11.4	1746	1 S19694	tenascin precursor
24	746.5	11.1	2019	1 J01322	tenascin precursor
25	744	11.0	1220	2 A56136	jagged protein pre
26	720.5	10.7	861	2 A48825	Notch homolog Mtc
27	717.5	10.6	1801	1 MMR5	laminin beta-2 cha
28	716	10.6	1306	1 S53868	laminin alpha-2 ch
29	709.5	10.5	1810	1 A32230	tenascin precursor

30	704.5	10.4	3084	1 MMSA	laminin alpha-1 ch
31	695.5	10.3	1790	1 MMR1	laminin beta-1 cha
32	692	10.3	3075	2 S14458	laminin alpha-1 ch
33	690	10.2	1408	2 S16148	gene serate prote
34	685.5	10.2	1798	2 S53869	laminin beta-2 cha
35	683.5	10.1	3635	2 T10053	laminin alpha 5 ch
36	676.5	10.0	3672	2 T32433	hypothetical prote
37	676.5	10.0	3704	2 T32316	probable laminin a
38	671	9.9	1722	2 E89753	protein F11C7.4 (1
39	667	9.9	1766	1 MMH01	laminin beta-1 cha
40	666	9.9	2823	2 T23064	hypothetical prote
41	666	9.9	2823	2 F87908	protein T22A3.8 (1
42	666	9.9	3102	2 T43291	laminin alpha cha1
43	664	9.8	1786	1 MMSB1	laminin beta-1 cha
44	657	9.7	1429	2 S06434	homeotic protein 1
45	641	9.5	1797	2 A55677	laminin beta-2 cha

ALIGNMENTS

RESULT 1

T13954

MEGF6 protein - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C:Accession: T13954

R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.

Genomics 51, 27-34, 1998

A:Title: Identification of high-molecular-weight proteins with multiple EGF-like mot.

A:Reference number: 214126; PMID:98360089; PMID:9693030

A:Accession: T13954

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1574 <MAX>

A:Cross-references: EMBL:AB011532; NID:93449293; PIDN:BA432462.1; PID:93449294

A:Experimental source: strain Sprague-Dawley; brain

C:Genetics:

A:Gene: MEGF6

Query Match	29.0%	Score 1958	DB 2	Length 1574
Best Local Similarity	41.3%	Pred. No. 1.1e-96		
Matches 344	Conservative 77	Mismatches 306	Indels 106	Gaps 16
QY	95	CP-GFYESGECVP--HCADKCVHGR-IAENTQCEFGMGTCSSACDGDHMGFICTS 150		
DB	602	CPKGFY--GKCRKRKCACNR--GRCHRLYGACICDPLGRCHLACPPMAFGPGCSE 656		
QY	151	RCQCKNG--ALCNPTTGACHCAAGFGMRCEDECGYGNDCRORCCOGATCDHYTG 208		
DB	657	DCLEQSHTRSCNPKDSCSCAKAFQGBRCABECSEGFEGCRHRCCTQPGVACDPVSG 716		
QY	209	ECR--CPFGYTGAFCEDLCPPGKHGPOCEORPCONGVCHHVTGEGSCPSGMGTVCGQ 266		
DB	717	ECRQCCPPGQGEDCGDECPVGTGVNCSGSCSV-GAPCHRYAGEICLCPKGTGEBOGA 775		
QY	267	PCPGRGRKNCSCQC--OCHNGCTDAATGQCHCSPGTGEGCQDECPYGTGYLCAETCQ 325		
DB	776	DCPGRWGLGQCEICPACCEHASCNPETGICLDPFGVGRSCDQYCSAGWGTGTCQIRCA 835		
QY	326	CVNCG--KCYH-----VSGACIC 341		
DB	836	CANDGHDPTTGRCSCAPGWTGLSCQACDSGHWGPPDIPHCNCSAGHCNDVAVSGCLIC 895		
QY	342	EAGFAGRCERARLCPGLYGIKCDKRCPCPLHENTHSCPMGSCACRPGMGLCNETCS 401		
DB	896	EAGFEGRCFCE--QSCRGQYGPSCQKRC--EHGAADHVSAGATCPAGMNGSFCBAACP 952		
QY	402	PGFYG-----EA-----CQIICSQNG 418		
DB	953	AGFGLDSDSACNCSAGAPCDAYTGSCICPAGRWGPRCAOSCPPLRTGLNCSQICFCNG 1012		
QY	419	ADDQSVTGKCTCAGPFGKIDQSTPCPLGTGYGINCSRGCKNDVAVCSPVDSQSCCKAGWH 478		

```

Db 1013 ASCDSTGCGCHCAPGMMGPTCLQACPPGLYGRKNCOSHCRCRNGGRDPTILAGCTCPREGMT 1072
Qy 479 GVDSCIRCPSTGTMGRCNLTCCOLINGACNTLDGTCTCAPGRNGEKECELPDQGTGLNC 538
Db 1073 GLACEHECLPRGHYAMCCOLNCSLHGGICDRLTGLCLCPAGMTGDCOSSVSGTFVNC 1132
Qy 539 ABRCDSCSHADGCHPTTGHCRCLPMSGVHDSVCAEGRMGPNCLPCYKNGASCSPDDG 598
Db 1133 EEHCAKRCASGHVHTGACFCPPGMWRPHCEQACPRGMFGEACAQCLCLPTNASHHTVG 1192
Qy 599 IECBAGFGTTCQRCISGFGYGRHSQTCPOCVHSSGGCHHTGLCDLDPFTALCNE 658
Db 1193 ECRCPGFGTGLSCQACQCTGKDCENLC-OCPEFTMACDPAISGVTCAAGHGTGLQ 1251
Qy 659 VCPSGFGRKNCAGICTCTNNGTNCPIDRSCQCYPGMIGSDCSOPCPRAHMGPNCTITNC 718
Db 1252 KCPSGYGRGCEHICLCLNGCTCDPATGACYCPAGFLGADCSLACQGRFSGCAHYAC 1311
Qy 719 HNGAFCSAYDSECKTPGWTGLYCTQRCPLGFGYKDCALICQCCQNGADCDHISGQCTRT 778
Db 1312 RCGAACDPVSGACISCPGKTGVRCEHGCQDPRFGKCELCACRNGGLCHATNGSCSPL 1371
Qy 779 GEMGRHCEKCPSGTYGYCROICLCLNSTCDHITGTCTCSPGMKGARGCDA 831
Db 1372 GMMGRHCEKCPAGRYGAACLECFQONNGSCPEPTTGACLGFGFYGACENS 1424

```

RESULT 2

727283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27283

R:Alnscough, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: 220336

A:Accession: T27283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1620 <WIL>

A:Cross-references: EMBL:AL110498; NID:el542303; PIDN:CAB54471.1; CESP:Y64G10A.f

A:Experimental source: clone Y64G10A

C:Genetics:

A:Gene: CESP:Y64G10A.f

A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match

Best Local Similarity 28.2%; Score 1900; DB 2; Length 1620;

Matches 331; Conservative 88; Mismatches 307; Indels 132; Gaps 17;

```

Qy 95 CP-GEYSEGEVCVPHCADKCVHGRCTAP-NTQCEPFGMGTCSSACDDHMGPHCTSRG 152
Db 712 CPDGFY--GSCNLCRCMCPNGRCDPVFGYCTCPDGLYGSCERKCPHPTGKNCRCPC 769
Qy 153 QC--KNGALCNPTTGACHCAAGFRGRCEDRCEQGTGNDCHORCCQNGATCDHVTGEC 210
Db 770 KCARENESECDEITGCKRCRPGYGHCKRMCSPGLFGAGCMAKCCSPAGIRCDPVTGDC 829
Qy 211 --RCPGYGACCEDCPRGKHGPOCEORPC-----QNGVCYHVTGECSPSGMGT 262
Db 830 TKKCRAGTGNLCLDPCRPAGTGYDEQKCSADVASPKSKYCHNVTCTCLPEKTRP 889
Qy 263 VCGQ-----PCPEGRFRKNCSECOCHNGTCDTAAATGCHCSPGYT 303
Db 890 LDDQCLIFETIEFDIAFSINVIACAPNTYGRPCAHTCGCVNCAKAKDESDGSHCRPGY 949
Qy 304 GERCDCEPVGYTGLCAFTCCCVNGKCYHVSAGCLCAGRAGERCEARLCPREGLYGK 363
Db 950 GATSCVCTPGRFGIDCMOLCKONGAICDTSNGSCCAPGWSKKCD--KACAPGTGKD 1008
Qy 364 CKRCPCHELNTHSCHPMGECACAKRMSGGLYCNETCSGFGFGEACQCLCSQNGADCS 423

```

```

Db 1009 CSKRCDC-ADGNH-CDPSDSEICCPGKGRHKRCDETCDSGLRGAGCGICSCQNGATCS 1066
Qy 424 YTGKTCAPGFGIDSTPCPLGTGYINCSRGC--KND-----AVCSPVDSCTC 473
Db 1067 VTGSCCECRGRMKKCKDRCPGRREBEGCNALCDCTTNDTSMYNPFAVARCHVTGECRC 1126
Qy 474 KAGMHGVDSCINCPGSGTWGFCNLTCQCLNGACNTLDGTCTCAPGRNGEKECELPQD-- 531
Db 1127 PAMTGPDCQTCPLRGHREGCRHSCQCSNGASCDRVTFCCPSGFMKNCSESECEGL 1186
Qy 532 -----CTGLNCAERDCSHADGCHPT 553
Db 1187 WGSNCKHCLHNGGECNENGDCEIDGTWTPSLCPFGQFRNCAQRNCRKNGASCDRK 1246
Qy 554 TGHCRCLPMSGVHCDVCAEGRMGPNCLPCYKNGASCSPDDGICEAPFRGTTCOR 613
Db 1247 TGRCECLPMSGVHCDVCAEGRMGPNCLPCYKNGASCSPDDGICEAPFRGTTCOR 1306
Qy 614 TSPGFYGRHSQTCPOCVHSSGGCHHTGLCDLDPFTALCNEVC----- 661
Db 1307 PCLKGYGRHCSQSC-RCANSKS-CDHISGRCCQPRGYAGHSTELCPDGTGSESCQKC 1364
Qy 662 -----GGRGKNCAGICTCTNNGTNCPIDRSCQ 690
Db 1365 DCGENSMCDALISGKCFKRGHSGSDKSGCVGRREPDCNQLSCBNGCVCSSTGSCVC 1424
Qy 691 YPGWIGSDSCOPCPRAHMGPNCTITNCNCHGAFCSAYDECKTPTGWTGLYCTQRCPLGF 750
Db 1425 PPGYIGTKCEINAOQSDRFPCTCEKICNGENGTGCDRLTGTQCHLDPFTMTQCNQVPEBR 1484
Qy 751 YKDCALICQCCQNGADCDHISGQCTCTRTGFMGRHCEQKCPSTGYGCAQICDCLNNTG 810
Db 1485 FAGCKEKKRCANG--HCNASSSECKNCNLGFTGSPCQSCPSGRKYGMCTLDECYQARNC 1543
Qy 811 DHITGTCVCPGKMGARGC 828
Db 1544 DPVQGCCDCPPGRYGRSNC 1561

```

RESULT 3

726972

hypothetical protein Y47H9C.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T26972

R:Harris, B.

submitted to the EMBL Data Library, October 1998

A:Reference number: 220293

A:Accession: T26972

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1111 <WIL>

A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4

A:Experimental source: clone Y47H9C

C:Genetics:

A:Gene: CESP:Y47H9C.4

A:Map position: 1

A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homol

Query Match

Best Local Similarity 26.8%; Score 1805.5; DB 2; Length 1111;

Matches 373; Conservative 162; Mismatches 423; Indels 215; Gaps 40;

```

Qy 21 GASPPLNLEDPNVCSHWEYSVTYVESYHPDPDIYI-----SCTILMMFKTRHR 73
Db 35 GTTEP---QGDHYCT-----VKTIVDDY--ELKVIHYVYVYDTEGCLNPLLGFQC---- 80
Qy 74 VSYRTAYRHGKTYRRK-----SQCCPFYSEGE--MCVPHCADKCVHGRCTAPNTC 124
Db 81 ----TYEKRGQASVYRQRLVKKKTYKKQCCDGYQTKDFCLPDCNPPCKKKKCIPIRGK 136
Qy 125 QCEPFGMGTCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCEDRCEQ 184

```


Db 623 NGYICPKGTTGVNCEKIDDCASNLCNKGKIDKIDGYECTCPGYTKLKNININEC 682
 QY 486 -----CPSGTWGFGC--- 495
 Db 663 DSNPCRNKGCTCKDOINGFTCYCPDGYHDHMCISEVNECNSNPCIGHACHDGVNGYCDCE 742
 QY 496 -----NLNCO---CLNGGACNTLDGT---CTCAPRGNGECEL-----PC-Q 530
 Db 743 AGHSGSNCDINNNECESNFCNMGCTCKDMTGATYICTCAGFSGPNCQTNINECSSNPLN 802
 QY 531 DGT-----YGLNC-----AERCD-----CSHADCHPT---TGHCRCLPGMS 564
 Db 803 HGTCIDVAGYKCNMFLPYTGALCEAVLAPCAGSPCKNGKNGKESDEFTFSCBEPBQM 862
 QY 565 GVHSD-----SYCABGRNGPNCSL-----PCYKNGASC 593
 Db 863 GQCEIDMNECVNPRNCRNGATCQNTNGSYKCNCRPGYTGRCNCEMDIDDCQNPNCNGSC 922
 QY 594 SPDDGI-----CECARGFNGTTCOR-----ICSPGYGHRG 624
 Db 923 S--DGINMFCNCFAGFPGKCEEDINECASNPCKNAGCTDCVNSTYCTCOPGFSGTHC 980
 QY 625 SQTCPQCVHSS-----GPCNHTTGL-----CDCLPFTGALC---NE----- 658
 Db 981 ESNTPDCTESSCFNGGTC--IDGINTFTCQCPGFTGSGYCOHDINECDKPCLANGTCQD 1038
 QY 659 -----VCPSPGRGKNGAGI-----CTCTNNNGC---NPIDSCOCYPCWISDSCOP 702
 Db 1039 SYGTCTKCTPGGYGLNCONLVNRCMDSSPCKNNGKRCWQTNFYR--CEKSSMTQVYCDVP 1097
 QY 703 ---CPRA--HWPNCIHTCNHNGAF--SAVDECKCTPGMTGLYCTORCLPFGYGRDC 755
 Db 1098 SVSEVAAKQGVDIYHL--CRNSGMCVDTGNTHTFCRCQAGYTSYCEBY-----DEC 1149
 QY 756 ALICCCONGADC-DHISG-OCTCRTGFMGRHCEOK----- 788
 Db 1150 S-PNPNCGATCTYLGYSCEVAGYGVNCSSEINECSLHPQNGGTCTIDLINTYKCS 1208
 QY 789 CPSPGYGGRQICD-----CLNNTSC--DHITG--TCYSPGKRGARCDAG 832
 Db 1209 CPRTQGVHCEINVDCTPTPDSLTLEPKCERNKGCIDRGGYNCICPFGVGRCE-- 1265
 QY 833 VTIIVNLNS--LSRTSTALPADS 853
 Db 1266 ----GDVNECLSN-----PCDS 1278

RESULT 5
 A40043
 Notch protein homolog TAN-1 precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change 02-Aug-2002
 C:Accession: A40043
 R:Ellisen, L.W.; Bird, J.; West, D.C.; Soreng, A.L.; Reynolds, T.C.; Smith, S.D.; Sklar, C.; Cell 66, 649-661, 1991
 A:Title: TAN-1, the human homolog of the Drosophila Notch gene, is broken by chromosomal
 A:Reference number: A40043; MID:91347367; PMID:1831692
 A:Accession: A40043
 A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
 A:molecule type: mRNA
 A:Residues: 1-2555 <ELN>
 A:Cross-references: GB:M73980
 C:Superfamily: notch protein; ankryrin repeat homology; EGF homology
 F:261-292/Domain: EGF homology <EGF1>
 F:994-525/Domain: EGF homology <EGF1>
 F:987-1018/Domain: EGF homology <EGF1>
 F:1149-1180/Domain: EGF homology <EGF2>
 F:1187-1218/Domain: EGF homology <EGF2>
 F:1233-1264/Domain: EGF homology <EGF3>
 F:1927-1959/Domain: ankryrin repeat homology <BEK3>
 F:1960-1992/Domain: ankryrin repeat homology <AN1>
 F:1994-2026/Domain: ankryrin repeat homology <AN2>
 F:2027-2059/Domain: ankryrin repeat homology <AN4>

F:2060-2092/Domain: ankryrin repeat homology <AN5>
 Query Match 15.3%; Score 1032; DB 2; Length 2555;
 Best Local Similarity 25.7%; Pred. No. 2,76-47;
 Matches 315; Conservative 84; Mismatches 304; Indels 522; Gaps 73;

QY 94 CCGFYSEGENCVPHCADKCYHGR-----IAPNTCCBEGWGTGSSACDDH 143
 Db 89 CALGF--SGPLCLPDLNACLTNRCNNGTCDDLTLLEYKRCRCPRGSGKSCQA----- 141
 QY 144 MGPHCTSRCCCKNGALCNPIIGA--CHCAAFGRGRCE--DRECG--TYGNDHQ- 193
 Db 142 --DPCASN--PCANGCCLPFEASTYICRPSFHGTCKQDVNEGQKPRCLRHGGTCNE 198
 QY 194 ---RC-----OQNGATC--DHYTGRCRPGYTAFC-- 222
 Db 199 VGSYRCVATHTGTPNCRERYVPCSPSCQNGGTCRPTGDTHTHCACLPRTGNGCEBN 258
 QY 223 DLCPFG--KHGPC-----EGRCP-----CQNGVCHVTG- 251
 Db 259 DDCPGNNCKNGACVDGVNTNCPREMTGQCTEDVDECOLPMNACQNGCTCKNTHG 318
 QY 252 -ECSPPSGMGTVCQ-----PCDEGRGRKNC--SQEC-- 281
 Db 319 YNCYCNVGMGTGEDCSENIDDCASACFRGATCHDRVASFTCEPHGRTGLCHLNDACIS 378
 QY 282 -QCHNGGTCA--ATGQ--CHCSPGYTGERCO--DECPVGTGVLAEITCQVNGGKCY 333
 Db 379 NPCNEGSNCDNPNVNGAKICTCPSGYTGPAQSQVDECSIGAN-----PCRHAGKCI 430
 QY 334 HVSA--CLCEAGNAGERCEARLCPBELYIGKCKRC--PCHLENTSHSPMSGE--CA 386
 Db 431 NTLGSFECQCLQGYTGPRCEIDV-----NECVSNPC--QNPATCIDQIGERQCM 477
 QY 387 CKPMSGLYC-----NE--TCSPGFYGEACQOI--CS-----CQ 416
 Db 478 CMPEYBEVHCYVNTDECASSPCLNKGKCLDKINFEQCEPRTGTHGLCDQVDECASTPK 537
 QY 417 NGADC-----DSV--TGKCTCAPRGKI 437
 Db 538 NGAKCLDGPNTYTCVTEGYTGTHCEVDIDECDDPCHYSGCKRGVATFTCLARPYTG 597
 QY 438 DCST-----PCPL-----GTGLNCS-----SRGCKADAVCSP 466
 Db 598 HCEFTNIECSSQPCRLKGTGCDPDNAYLCFLKGTTPNCEINLDCASSPCDSTGLDK 657
 QY 467 VDG--SCYCKAGMHGVDCSIR-----CPSGTGFEGCNL-----TC----- 499
 Db 658 IDGYEACAEPRGYTSGMNSNINDECAGNPCNNGGTCDEGNGFTCRCEGYHDPCLSEVN 717
 QY 500 ----OCLNGGACNTLDG--TCTCAPWRGKCEL-----P 528
 Db 718 ECNSNPVHACRDLNKGKCDXDPGWSGTNCDINNNECSNPNVNGGTCKDKMTSGIVCT 777
 QY 529 CQGTLYLNAERCD-----CSHADG--HTTGH--CRCLPGMSGVCHDSV--CAEG-- 575
 Db 778 CRGFSFSPNCOITNINECASNPCNLKGTCTDVAGYKCNCLPYTGACEVVALPACAPSPC 837
 QY 576 RMPNRC-----SLPYC-----KNGASCSPDQICACAPFGTGCQRI-----CSP 617
 Db 838 RNGEGRQSDYSEFSCVCTPAAGKGTCEVDINECVLSCRHGASQANTHGAYRCHCQA 897
 QY 618 GYFGHRCSTCPQCVHSSGPCHH-----ITGLDCLPFTGALCNE----- 658
 Db 898 GYSGRNCETDIDDC--RPNCHNGGSGCTDINTAFCDCLGFGFTGTCEDDINECASDPCR 955
 QY 659 -----VCPSGRFGKNCAG--ICT--CTNNGTNPIDR-----SCQCTPGW 694
 Db 956 NGANCYDCVDSYCTCPAGFSIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPPGF 1013
 QY 695 IGSDC-----SOP-----CPRAHMPNRC-----IHCN-----CHNGAF 723
 Db 1014 TGSYCOHYVNECDSPCLLAGTCTQDGRGLHRTCTPGCTGTPNCONLVNRCMDSSPCKNKGK 1073


```

Oy 699 C-----SOP-----CPRAHMPNC---IHTCN---CHNGAFC--- 724
Db 1018 CQVYNEDCSRPCLHGTGTCODSYGTYKCYCPQGYTGLNQNLYVRMDSAPCKNGRCMQT 1077
Oy 725 -SAYDECKCTPGMTGLYC---TORCPLEFY---GRKCALICO----- 760
Db 1078 NTQY--HCECRSGMTGVNCDVLSVSCVANAQKRGIDVTLICQHGGLCVDEGKHCHCOA 1135
Oy 761 -----CONGADC-DHISG-OCTCRGFMGRHCEOK----- 788
Db 1136 GYTGSYCEDEVEDCSNPNCQNGATCTDYLGFSCCKVAGYHGSNCSEIEINCLSQPCQNG 1195
Oy 789 -----CPSGTYGCGROI-----CLNSTC-DHITG--TCYC 819
Db 1196 GTCIDLTNYSKCSCPRGTOGVHCEINVDCHPRLDPASRPRCFNNGTCVDVGGYTCYC 1255
Oy 820 SPGMKARCDDQAGVIIVGNLN 840
Db 1256 PGFVGERCE-----GDVN 1269

```

RESULT 7

S1818
 notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S1818
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 113, 199-205, 1991
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383; PMID:1764995
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <MEI>
 A:Cross-references: EMBL:X57405; NID:957634; PID:957635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF2>
 F:1233-1264/Domain: EGF homology <EGF2>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 15.2% Score 1024; DB 2; Length 2531;
 Best Local Similarity 25.8%; Pred. No. 7.1e-47;
 Matches 315; Conservative 83; Mismatches 286; Indels 536; Gaps 74;

```

Y 86 TMYRRKSQCPCPFYSEGENCV---PHCADKCVH--GRCT---APNTQCCEPFGNGTNCSS- 137
Db 121 TLTEYKCRCPGWM--SGKSCQADPCASNPNCANNGOCLPFESSYICGCPFGHGPICRQD 178
Oy 138 -----ACDGDHMGPHC-----TSRCCCKNALCNP--- 162
Db 179 VNECSQNPGLCHNGCTCHNEISYRCACNATHTGPHCELPIYPCSPSPQNGTCTCAPICD 238
Oy 163 ITGACHCAAGFRGWRCEDECRCEOGTYGNDCHQRCQCONGATC--DHY--TGECRCRPGYTGA 220
Db 239 TTHCACLPGFAGNCEENV-----DCPGN--NCKNGACVADVNTYCRCPREPTGQY 291
Oy 221 C-EDLCPRGKHGRCQRCR--CQNGVCHAVTG--ECSGPRGMMGYVCGQ----- 266
Db 292 CTEDV-----DEC-QLMPNACQNAAGTCHNSHGVCNVCVAGWTGECDSIDINDCASNA 343
Oy 267 -----PDEGRFGKNC--SQEC-----QCHNGTCA--ATGQ--CHCSFG 301
Db 344 CQCGATCHDRVASFYCECPHGRTGLLCHINDACISNPEGSSNCDPNVNGAICTCPRG 403
Oy 302 YTGRCQ---DECPTYGVGLCAETCCQCVNGSKCYHVSQA--CLCEAGFAGRCERARLCP 356
Db 404 YTGPRACSDVDCAALGAN-----PCBHAKCLNTLGSFPCQCLQGTGPRCEIDV-- 453

```

```

Oy 357 EGLYIKCDKRC---PCHLENTSCHPMGSE--CAKPRGWSGLY----- 396
Db 454 -----NECISNPC--QNDATCLDQIEFGQICIMGEYGEYCEIINTDECASPCLHN 502
Oy 397 -----NE--TCSAPFYGAQO---QICS---CONGADC--DBVTCKTCAPFGKID 438
Db 503 GRCVDRKINEFLCQCKRGSGLHCOYDYDECACTPCKRNGAKCLDGENITYTCTBEGYTH 562
Oy 439 CST-----PCPLGTGYNCSRCGCKNDKAVCSPYVDSCTCKAGW----- 477
Db 563 CEVDIDCEDPDPCHIGL-----CK-DEVAR---FTCLCQPGYGHHEFTINBCH 608
Oy 478 -----HGVDCSTR-----CPSGTWGRCNLTQ-----CLNGACNTLDG--TCTAP 518
Db 609 SPCRHHGTCODRDNYLYLCLKGTTPNCEINLDCASNPDSCCTCKIDYEACAP 668
Oy 519 GWRGKCEL-----PCODGTGYNLCARCDCHADCHPTT----- 554
Db 669 GTTSMCNVNIIDECAGPSCHNGTCEBDGIAGFTC--RC---PBGYHPTCLSEVNECS 722
Oy 555 -----GHCR-----CLPGMSGVHCD-----SVCAEG 575
Db 723 NPCIHGACRDGLNGYKCDAPGMSGTNCDINNNECESNPVNGTCKDMTSGVYTCREG 782
Oy 576 RMGPNC-----SLPCT-----CKNGA 591
Db 783 FSGPNCOTNINECANPNCLNCGTCLIDVAGYKCNCPLEPYGATCEVILAPCATSPCKNSG 842
Oy 592 SCSPDGI-----CEGAPGRGTQO-----RCSGPFYGH 622
Db 843 VKESIEDYESSCVCPPTGOGGTCEIDINECVKSCRHGASQONTNGSTKLCQAGYGR 902
Oy 623 RCSQTPQCVRHSSGCPCH-----ITGLCDLPGFTGALCNE----- 658
Db 903 NCESDIDC--RPNCHNGSGCTDGVNAFAFCDLPGFQGAFCEDINECATNPONGANC 960
Oy 659 -----VCSGRPRKNCAG--ICT---CTNNGTNPIDR-----SCQCTPRNGISPC 699
Db 961 TDCVDSYTCCTPTGPNHICENTTPDCTESSCFNNGTCT--VDGINSFTCLCPRGFTGYSY 1018
Oy 700 -----SOP-----CPRAHMPNC---IHTCN---CHNGAFC--- 724
Db 1019 QYDVNECBSRPLHGTGTCODSYGTYKCYCPQGYTGLNQNLYVRMDSAPCKNGRCMQT 1078
Oy 725 SAYDECKCTPGMTGLYC---TORCPLEFY---GRKCALICO----- 760
Db 1079 TQY--HCECRSGMTGFNCVLSVSCVANAQKRGIDVTLICQHGGLCVDEEDKHCHCOAG 1136
Oy 761 -----CONGADC-DHISG-OCTCRGFMGRHCEOK----- 788
Db 1137 YTGSYCEDEVEDCSNPNCQNGATCTDYLGFSCCKVAGYHGSNCSEIEINCLSQPCQNG 1196
Oy 789 -----CPSGTYGCGROI-----CLNSTC-DHITG--TCYCS 820
Db 1197 TCIDLTNYSKCSCPRGTOGVHCEINVDCHPRLDPASRPRCFNNGTCVDVGGYTCYC 1256
Oy 821 PGWKARCDDQAGVIIVGNLN 840
Db 1257 PGFVGERCE-----GDVN 1269

```

RESULT 8

A49128
 cell-fate determining gene Notch2 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
 C:Accession: A49128
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 Development 116, 931-941, 1992
 A:Title: Notch2: a second mammalian Notch gene.
 A:Reference number: A49128; MUID:93202015; PMID:1295745
 A:Accession: A49128
 A>Status: preliminary; not compared with conceptual translation

A: Molecule type: mRNA
 A: Residues: 1-2471 <ME1>
 A: Experimental source: Schwann cell
 A: Note: Sequence extracted from NCBI backbone (NCBI:127811)
 C: Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F: 264-995/Domain: EGF homology <EGX1>
 F: 799-830/Domain: EGF homology <EGF1>
 F: 877-908/Domain: EGF homology <EGX2>
 F: 1029-1060/Domain: EGF homology <EGF>
 F: 1067-1098/Domain: EGF homology <EGX3>
 F: 1153-1184/Domain: EGF homology <EGF3>
 F: 1191-1222/Domain: EGF homology <EGX4>
 F: 1876-1908/Domain: ankyrin repeat homology <AN2>
 F: 1809-1941/Domain: ankyrin repeat homology <AN1>
 F: 1943-1975/Domain: ankyrin repeat homology <AN3>
 F: 1976-2008/Domain: ankyrin repeat homology <AN4>
 F: 2009-2041/Domain: ankyrin repeat homology <AN5>

Query Match 14.88; Score 998; DB 2; Length 2471;
 Best Local Similarity 24.48; Pred. No. 1,7e-45;
 Matches 321; Conservative 79; Mismatches 322; Indels 596; Gaps 70;

17 CHMIGTASPLNLEDPNCSHESVSVTQESY-----PHFPDIYTSCTDIL 64
 121 CHMIS-----WDFTCTQVGFTEGKCCOMTDVCLSHPCEN--GSTCSSVA 163
 65 NMFCTHRVRYRATYRNGEKTMRKRSQCCPGFSESEMCVPRCAKCVHGRCLAPRTC 124
 164 NQFSC-----RCPAGI--TGQKCDADINEDIDRGKOHGGTC 198
 125 QCEBGMGTNCSSACDDDHMPHCTS-----RCQCKNGALC--NPITGACHCAAGFRG 175
 199 LNFLE--GSYRQ--CPQRFTEQHCDSPYPCAPSPVNGGRCRTGDTSTCHCLPREG 254
 176 WRCDRECEQGTGMDCHORCOQONGATC-DHY--TECEPCRPYTGAFNC-EDLCPRGKHG 232
 255 SNCRNTD-----DCPNH--KQNGVGVCDVNTYNCRCRPMQMGFCTEDV-----D 300
 233 QC-EORPCONGVGVCHVHTG--ECSCPGMMGTVCQGP----- 267
 301 ECLLDPRNACONGGCTCTNRNGCYCYVNGMSGDDCSEHIDDCARASCTPGSTCIDRVASF 360
 268 ---CPGRFGRKNC--SQEC--OCHNGTCDG--ATGQ--CHCSFGYGERGQ--DECP 312
 361 SCLEPEGAGGLCHLDACISNPKHKGALDPTNLNGYITCQAKGACSTEDVDECA 420
 313 VGTGVLCAETCCQCVNGSKCHVSGA--CLCEAFAGERCE----- 351
 421 M-----ANSNPEHAGKCVNTDGAFFHCCLKYGAGRCEMDINECHSDPCQNDATCLD 473
 352 ---ARLCPREGLYGIKDKR-----CP-----CHLE- 373
 474 KIGFTCLCMRFGKGVCELEVNRCQSNPCVNGQCVDKVNRFCOLCPRTGTYCQIDTI 533
 374 ---NTHSC--HPMSGECACKPGWSGLYCNET----- 399
 534 DDCSFTPLNCAKCIDHNGYEGCAGFTGLCDENIDNDRPHHGGQDQDIDSYTC 593
 400 -CSGCFEAGC--QOICSC-----QNGADCSVTG--KCTAPRGKIGDC-----STP 442
 594 ICNNGYGAICSDQIDECYSPCLNDGRCIDLVNGYOCNCQPCPGTSGLNCEINFDDCASNP 653
 443 CPLGTY--GIN-----SRG--CKNDAYC----- 464
 654 CLMGACVADGIRYSCVSPGFTGRCNIDIDECASNPCRDATCINDVNGFRMCPCREPH 713
 465 ---SP--VDGSCT-----CKAGHGVDCSI----- 484
 714 HPSCQVNECLSPCLHGNCTGLSGYKCLDAGWGINCEYVKNNECLNPNCONGCTCN 773
 485 ---KCPSGTWGFGCNLTQ-----CLNGAC----- 507
 774 NLVNGYRCTCKKGFKNYCNQVNIIDECASNPLNGTCLDVSgyTCHCMPLPYTGKNCQTV 833

508 -----NTLDGTCAPRGWGEKELPCQDGTGYGLNCAERDCSHAD 548
 834 LAPSPNPECNAAYCKEAPNESEFTCLAPWQORCVDDV-----CVSK--PCNNNG 886
 549 GCHPPTTG--CRCLPGWSGVHCDSDVCAEGRFGPNCSLPYCKNGASC--SPDDGICECAP 604
 887 ICHHTQGSYNKECPRGSGMCEE-----DINDCLANPCONGSCYDKKNTSPCLCLP 939
 605 GFRTTCQR-----ICSPFYGHRCSTQPCQVHSS----- 635
 940 GFVGDCKQTDNNECLSEPCKNKGTCSDPVNSYTCPCPAGFHVGCENNIDCTESSCPNG 999
 636 GRCHHTIGL-----DDCLPFGTALC-----NE-----VPPSGRFG 666
 1000 GTC--VQINSFSLCPVGFTEGFPCLDLINDINCSSNPLCNSGTQVVDGLTGYCTCTPLGYTG 1057
 667 KNC---AGICT--CTNNGTC--NPIDRSCQCYEGWIGSDSQ----- 701
 1058 KNCQTLVNLCSPPCKKNGCTAQEKARPCLPFGWGAAYCDVNLVSCKAALQKVPYE 1117
 702 -----PCPPAHMGPNC--HTC--NCHNGAFCSAIDG--ECKCTP 735
 1118 HLCQHSGLICINAGNTHHCQCPFLGYTGSYCEQDLDECASNPQHGATGSDFTGYRCCEVP 1177
 736 GMTGLYCGNR-----CPGFGV-----KDCALICQON 763
 1178 GYGVCNCEYEVDQCONPCONGGTCIDLNVHFKSCSPGTRGLCEENIDNCAAPPLN 1237
 764 GADC-DHISG-----OCTCRTEGM 781
 1238 GGQCVDRIGYGSCHCLPGFAGERCEGIDNECLSNPCSEGLDQIKNNKQCVCRSAFT 1297
 782 GRHCE---QKCPSTYTGRCQICDCLANSTCDHITGT-----CYSGMGKAGACDA 831
 1298 GRHCETFLDVCPRK-----PCJNGTCVAAVNPDPGFCRCPGFGSGARCOSS 1345

RESULT 9
 S42612
 Transmembrane protein precursor - zebra fish
 C: Species: Brachydanio rerio (zebra fish)
 C: Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 02-Aug-2002
 C: Accession: S42612
 R: Bierkamp, C.; Campos-Ortega, J.A.
 Mech. Dev. 43, 87-100, 1993
 A: Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patte:
 A: Reference number: S42612; MUID: 94128602; PMID: 8297791
 A: Accession: S42612
 A: Status: preliminary
 A: Molecule type: mRNA
 A: Residues: 1-2437 <BIE>
 A: Cross-references: EMBL: X69088; NID: g433866; PID: CAA48831.1; PID: g433867
 C: Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F: 755-186/Domain: EGF homology <EGF1>
 F: 1023-1054/Domain: EGF homology <EGF2>
 F: 1185-1216/Domain: EGF homology <EGF3>
 F: 1315-1947/Domain: ankyrin repeat homology <AN1>
 F: 1948-1980/Domain: ankyrin repeat homology <AN2>
 F: 1982-2014/Domain: ankyrin repeat homology <AN3>
 F: 2015-2047/Domain: ankyrin repeat homology <AN4>
 F: 2048-2080/Domain: ankyrin repeat homology <AN5>

Query Match 14.68; Score 987; DB 2; Length 2437;
 Best Local Similarity 24.88; Pred. No. 6.3e-45;
 Matches 310; Conservative 81; Mismatches 320; Indels 538; Gaps 70;

91 KSQCCPFYSGEMCVHCHADKCVHRC-----IAPNTQCCPRGNGTNCSSACD 140
 85 KCDCVLSG--SDRLCLPVRNACNNSPCRNCGTCSLLTLTFTCTCRCPGNSGKTQQLA-- 140
 141 GDHWGPRCTSRCCKNALCNPTIG--ACHCAAGFRMGREDRCQGTGNDCH-QRCQC 197

```

Db 141 -----DPCASN-PCANGGCSAFESHYICTCPENFGOTCRDQV-----NECAVSPSPC 188
Qy 198 QNGATCDHYTGE--CRCPGTGAFCEDL-----CPGPK 229
Db 189 RRGGTCINVGSLCRCPPETGPHQRLXORCLSPCSGGTCVTSPTHTSCLPGRF 248
Qy 230 HGPQCE-----ORC-----PCQNG 243
Db 249 TGOTCEHNDYDCTQHACENGSPCIDINTYNCHCDKHTGVCCTEDVDECELSPNACQNG 308
Qy 244 GYCHHTG--ECSGPGMAGTYCGO-----PCEGREGKN 276
Db 309 GTCCHNTGGFHCYCVAGWGDDCSSEIDDCASAACSHGATCDHRVASFECEPHGTGLL 368
Qy 277 C--SQEC--OCHNGGTCA--ATGQ--CHSPGYTGRCQ--DECPVGTGVLCAETC 324
Db 369 CHLDACISNPPCKGKSCNCTPNVSGAICTCPGTGSGACNDIDECISGAN-----420
Qy 325 QCVNGSKYHVSQA--CLCEAGFAGEARCLPEGLYIKDKRCPCHEMTSCHPNS 382
Db 421 PCEHGGRCINTKSGSFQCKLQGYEGPRCEMDV-----NEC-KSNPC--QNDATCLDQI 470
Qy 383 G--EACRKGMSGLVCNER-----CSRGFGEACQ--QIC 413
Db 471 GGFHCICMFGYEGVFCQINSDDCASOPCLNGKCIDKINSFCECPKRGFSGLQVDVDE 530
Qy 414 S---CONGADCDVYTK--CTCAPGFKGIDC-----STPCPL-----GTYGINCSSR 455
Db 531 ASTPCKNGAKCTDGPKNKYCECTPGFSGIHCELDINECASSPCHGVCHDVAASFCDGR 590
Qy 456 CG-----CKDAVCSPIVDS--CTCKAGMHGVDCSIR-----485
Db 591 PGYTRGLCTNINECLSOPCRNGGTQDRENAVICTCPGTGVCNEINIDCKRRKPCDY 650
Qy 486 -----CPSGTGFSCNLTQO-----CLNGACNTLDG-----TCGAPGMRG 522
Db 651 GKCIDKINGEVCERGYSKMCNINIDCALNPGCHNGTGC--IDGVNFTCLCPRGFRD 708
Qy 523 EKC-----ELPCQDGYTLNC--AERC--DCSHADGCHP 552
Db 709 ATCLSQHNECSSNPCHIGSCLDQINSYRCVCEAGWMGRMCDININCLSNPCVNGSTCD 768
Qy 553 -TGH--CRCLPGMSGVHCDVCAEGRMGP-----NCSL-----583
Db 769 MTSGLTCTCRAPFSGPNCCMINIECASNPCLNMGSCIDVAGFKCMPLPYGECENVL 828
Qy 584 -PCY--CANGASCSPDD-----GICE-----601
Db 829 ACSRPCKNGVGNESDEFSQSCNCPAGMGTCEVDINECVNRNPGTNGVCEMLRG 888
Qy 602 ---CAPGRTTCOR-----ICSPGFGHRCSTQCPQCV 632
Db 889 FOCRNPFGTGTALCEDDIDCEPNPCSNGVQDVRNGFVCLAGFRGERCAEDIDECV 948
Qy 633 HSSGPHHTITGLCDLPGFTGALCNEVCSGRFKNC--AICT--CTNNGTCNPIR 686
Db 949 --SACCRKNGNCTDCVNST--CS--CPAGFSGINCEINTPDCIESSCFNGTGC--VGG 999
Qy 687 ---SCOCYPGWIGSDC-----SQP-----CPRHMPNC--IH 714
Db 1000 ISSFCVCLPFTGNYCQHDVNECDSPQNGSGQDGYTKCPCPHRYTGLNQSGLVR 1059
Qy 715 TCN---CHNGAFC--SAVDECKTPGWTGLYCTQ-----744
Db 1060 WDDSPCKRNGSGSMQOGASFTQCSAGWTGYCDVPSVCEVAARQGVSAVLCHRAQ 1119
Qy 745 -----RCPLGFYKDC--ALIQ--CONGADC-DHISG--QCICRGRFMRHBE 786
Db 1120 CVDAGNTHLCRCQAGYTSYCOEQVDECOQPNPCQNGATCTDGLGYSCECBVGHQMNCS 1179
Qy 787 OK-----CPSGTYGGRQICD-----CLNN 807
Db 1180 KEINIELSOPQONGSTCIDLVNTYKSCSPRGTOGVHCEIDIDDCSPSVDPLTGEPRCNG 1239

```

```

Qy 808 SNC-DHITG--TCYCSPPMKARCDQAGVITVGNLNSLSTALPADSY 854
Db 1240 GRCVDRVSGYCVCCPAGFGERCE-----GDVNE-CLSDPCDPSGST 1280

```

RESULT 10

A24420

notch protein - fruit fly (Drosophila melanogaster)

N.Alternate names: neurogenic repetitive locus protein

C.Species: Drosophila melanogaster

C.Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

C.Accession: A24420; A24768; S09358; A05267

R.Kidd, S.; Kelley, M.R.; Young, M.W.

Mol. Cell. Biol. 6: 3094-3108, 1986

A.Reference number: A24420; MUID:87064624; PMID:3097517

A.Accession: A24420

A.Molecule type: DNA

A.Residues: 1-2703 <KID>

A.Cross-references: GB:K03508; NID:9157991; PIDN:AAA28725.1; PID:9157993

R.Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.

Cell 43, 567-581, 1985

A.Reference number: A24768; MUID:86079539; PMID:3935325

A.Accession: A24768

A.Molecule type: mRNA

A.Residues: 1-48, 'I', '50-118, 'R', '120-230, 'I', '232-256, 'N', '258-266, 'A', '268-872, 'R', '874-9

A.Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2

R.Nautz, D.

Nucleic Acids Res. 17, 6463-6471, 1989

A.Title: Hypervariability of simple sequences as a general source for polymorphic DNA

A.Reference number: S09358; MUID:89385974; PMID:2780284

A.Accession: S09358

A.Molecule type: DNA

A.Residues: 2505-2551, '0000', '2552-2576, 'E', '2578-2604 <TAU>

R.Wharton, K.A.; Vedvodnick, B.; Flinerty, V.G.; Artavanis-Tsakonas, S.

Cell 40, 55-62, 1985

A.Title: opa: a novel family of transcribed repeats shared by the Notch locus and oth

A.Reference number: A05267; MUID:85099329; PMID:2981631

A.Accession: A05267

A.Molecule type: DNA

A.Residues: 2504-2576, 'E', '2578-2611 <MHA>

C.Genetics:

A.Gene: notch; opa

A.Cross-references: FlyBase:FBgn0004647

A.Map position: 8,96-9,36

A.Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3

C.Superfamily: notch protein; ankyrin repeat homology; EGF homology

C.Keywords: differential; tandem repeat; transmembrane protein

F:27-43/Domain: transmembrane #status predicted <TMN1>

F:297-328/Domain: EGF homology <EGX1>

F:530-561/Domain: EGF homology <EGF1>

F:568-599/Domain: EGF homology <EGF>

F:988-1019/Domain: EGF homology <EGX2>

F:1064-1095/Domain: EGF homology <EGF3>

F:1187-1218/Domain: EGF homology <EGX3>

F:1746-1762/Domain: transmembrane #status predicted <TMN2>

F:1950-1982/Domain: ankyrin repeat homology <AN1>

F:1983-2015/Domain: ankyrin repeat homology <AN2>

F:1988-2004/Domain: transmembrane #status predicted <TMN3>

F:2017-2049/Domain: ankyrin repeat homology <AN3>

F:2050-2082/Domain: ankyrin repeat homology <AN4>

F:2083-2115/Domain: ankyrin repeat homology <AN5>

F:2538-2568/Region: glutamine-rich

F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match

14.5%; Score 978.5; DB 1; Length 2703;

Best Local Similarity 26.8%; Pred. No. 2e-44;

Matches 290; Conservative 102; Mismatches 297; Indels 395; Gaps 70;

```

Qy 7 SCL-----SPICLLCHWIGTASPLNLED--PNVCSHMEISYTVQESYHPFDQIYITSC 60
Db 502 SCLDDPGFRFCYCMGFGFTGCEIDIDECQSNPC-----LNDGTC 541

```



```

OY 484 -----IRCPSSWGRGNCNLTC-QCLNG-----GACNTLDG-----TTCAPGRWG 522
Db 595 GSKCLDLYDKYLRCRCPPTGTVNCEVNIIDDCASNPCTFVCR--DGINRYDCVCOPGFTG 652
OY 523 ENCEL-----PCODGTYYGLNC-----AERCDCSHADGCH 551
Db 653 PLCANVEINCCASSPCGEGSSCVDSGNGRFLCPLPSLPLCLPANHPCNHRKCSHG-VCH 711
OY 552 PTTG--HCRCLPGMSGVHCDVSCABGRWGPCNSLPCYCKNGASCSPDDGI---CECAPGF 606
Db 712 DAGGFRVCYCEPWSGPRCSQSLA-----PDACESQPCOAGGTC--SDIGFRICAPGF 765
OY 607 RCTTCORT-----CSPGFYGHNCSTQPCQVHSS-----G 636
Db 766 GCHQCEVSLPCTPSLCEHGHCESPDRLLTYVCSPPGMOGPRCQDDVDEACAGASCPCGPRG 825
OY 637 PCHNITG--LCLDCLPFGFALCNE-----VCPSGRFGKNC 669
Db 826 TCTNLPGNPRICHRGTYGPPCDODIDODPNRCLHGSSCQDGVGFSCLDGFAGPRC 885
OY 670 A-----GICF-----CTNNGTCNPIDR 686
Db 886 ARVDDECLSSPCGPCTCTDHVASFYCACPRGYGFHCEIDLPCDSCSPSCFNCGTC--VDG 943
OY 687 -----SCOCYPMWISDC-----SOP-----CPRAHMGPNCIHTCN----- 717
Db 944 VSSFSCLCRPGYTGTHCYEADPCRSRPLCHGEGICNPTHPGEC--TCREGTSGSQCNP 1001
OY 718 -----CHNGAFCSAYDGECKTPGWTG-----LYCTOR----- 745
Db 1002 VDMCSQAPCQNGRCVYGYATCYICPPGWSRLCDIQLPCTEAAAMQVRLBQLQDEGK 1061
OY 746 -----CPGLGFYKDC-----ALICQCONGACCDHISG--OCTCRTEGMRICE 786
Db 1062 CIDGRSHYCVCPESRGTSHCEHEVDPCTAQCQHGTCRGYMGGYVCECPAGYAGDSCE 1121
OY 787 Q-----KCPSGTYGYG--RQICD-----CLNNS 808
Db 1122 DNIDECASQPCNGSGCIDLVARYLCSPPSTGLVLCLEINEDDCDLGFLDSGVQCLHNG 1181
OY 809 TCDHITG--TCYSPGKMGARCD 829
Db 1182 TCVDLVGGFRNCNCPGTYGLHCE 1204

```

RESULT 12

notch3 protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 02-Aug-2002

C:Accession: S78549; S71825

R:Joutel, A.; Tournier-Lasserre, E. submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <JOU2>

A:Cross-references: EMBL:U97669; NID:g2668591; PIDN:AB91371.1; PID:g2668592

R:Joutel, A.; Coppechot, C.; Ducros, A.; Vahedi, K.; Chabrial, H.; Mouton, P.; Alamowicz, M.M.; Weissbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MUID:97032728; PMID:8878478

A:Accession: S71825

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333,'G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2>

A:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

A:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

```

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology; laminin-type EGF
C:Keywords: tandem repeat; transmembrane protein
F:123-155/Domain: EGF homology <EGX1>
F:162-194/Domain: EGF homology <EGX1>
F:240-271/Domain: EGF homology <EGX2>
F:318-349/Domain: EGF homology <EGF>
F:473-504/Domain: EGF homology <EGX3>
F:853-884/Domain: EGF homology <EGF3>
F:928-959/Domain: EGF homology <EGX4>
F:1070-1126/Domain: laminin-type EGF-like homology <LEG>
F:1838-1870/Domain: ankyrin repeat homology <AN1>
F:1871-1903/Domain: ankyrin repeat homology <AN2>
F:1905-1937/Domain: ankyrin repeat homology <AN3>
F:1938-1970/Domain: ankyrin repeat homology <AN4>
F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 14.4%; Score 974; DB 2; Length 2321;
Best Local Similarity 25.3%; Pred. No. 3e-44;
Matches 304; Conservative 69; Mismatches 313; Indels 514; Gaps 68;

OY 107 PHCAD--KCVH-GRCI-APN---TCQCEPBGWGTNCSSACDGDHMGPHCTSRCKNGAL 159
Db 41 PPLDGSPCANCGRGTOLPRLBACLCPPGVYGERQLE-DPCHSGP-CAGRGVCQSSV 98
OY 160 CNPITGACHCAGFERGWRCE--DRCEQGTYGNDCHQRCQCONGATYCHVTGE-----CRC 212
Db 99 AGTARFSCRCRGRFGPDCSLPDCLS-----PCAHGANCs--VBPDRFLCSC 146
OY 213 PRGYTAGCE---DLCPFGKHGPOCEQRCPCONGVCHHTYG--ECSPSGMKTYGCP 267
Db 147 PPGTQNSCRSDVDCEKRGV-----PCRHGCTCNTGSRFCQCPAATYGPLCBNP 197
OY 268 CPEGFRKNCSEOCCHNGTCDAA--TGQCHSPGYTGERCQ--DECP----- 312
Db 198 AVP-----CAPS-PCRRGGRGSGDLTYACLPREGGNCENYNDCCGHNCINGT 250
OY 313 ---VGYTYGLC-----AETCQ-----CVNGKCYHVS--ACLCENGEFAGER 349
Db 251 CVDDGVNTYNCQCPPEMTGQFCTEDVDECOLPNACIHNGGTCTFNTLGHGSHCVANGWGES 310
OY 350 CEARL-----CPGLYGIKC--DKRC--PCLEMTHTSHCP 380
Db 311 CSQINDCATAVCFRGATCDRVASFYCACPMKGTGLCHLDACVSNPCHEDIAICDTNP 370
OY 381 MSGE--CAKPRMSGLYCNE----- 398
Db 371 VNGRAICTCPGFTGAGACDDQDVDECSIGANPCHELRGVNTQGSFLCQCGGTYGPRCT 430
OY 399 -----TSPGYGERCQ-----QICSONGADC-DSVTG 426
Db 431 DVNECLSGPCRNQATCLDRIGQFTICIMAGFTGYCEVDIDECOSSPCVNGVCYKDRVNG 490
OY 427 -KCTCAGFKGIDC-----STPCPLGTV-----GINCSSRCG-----C-KNDAY 463
Db 491 FSCCTPSPGFSSTQLDVBECASITPCRNAGKACVQDPGYEC--RCABGFSTGLCDRVNDD 548
OY 464 CSP-----VDG-----SCTCKAGMHGVDCS----- 486
Db 549 CSPDPCRRHRCVDDIASFSCACAGYGTGTCESEGVDECSRQPCRHGKCLDLYDKYLCRC 608
OY 487 PSCTGWGSCNLTG-QCLNG-----GACNTLDG-----TTCAPGRWGEKCEL----- 527
Db 609 PSGITTYGNCEVNIIDDCASNPCTFVCR--DGINRYDCVCOPGFTGPLCANVEINCCASSPC 666
OY 528 -----PCODGTYYGLNC-----AERC-----DCSHADGCHPTTG--HCRCLPGWS 564
Db 667 GEGGSCVDGNGRFLCPLPSLPLCLPSPHCAHPCSHGT-CYDAPGGRVCCEGWS 725
OY 565 GVHCDs-----VCAGRWGPNCSL--PCY-----CKNGAS 592
Db 726 GPRCSQSLADACESQPCRAAGTCSSDGMGFHCTCPRGVGRGCELLSPCTPNCEHGR 785
OY 593 CSPDDG--ICECAPGFRGTTCOR-----ICSPGFYGH 623

```


A:Cross-references: GB:X17530; NID:g10225; PID:g667061
 A:Accession: B40136
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 181-251,329-370,'R',372-408,'KA',411-441 <DE>
 A:Accession: C40136
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 'K',747-821,898-978 <DE>
 R:Hursh, D.A.; Andrews, M.E.; Raff, R.A.
 Science 237, 1487-1490, 1987
 A:Title: A sea urchin gene encodes a polypeptide homologous to epidermal growth factor.
 A:Reference number: A23116; MUID:87319677; PMID:3498216
 A:Accession: A23116
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'S',280-481,786-1064 <HUB>
 A:Cross-references: GB:M17421; NID:g161474; PIDN:AAA30050.1; PID:g552260
 R:Hunt, L.T.; Barker, W.C.
 FASEB J. 3, 1760-1764, 1989
 A:Title: Avidin-like domain in an epidermal growth factor homolog from a sea urchin.
 A:Reference number: A43131; MUID:89196806; PMID:2784773
 A:Contents: annotation
 C:Comment: EGF homology repeats 10-17 are spliced out in the short form (fibropellin Ib)
 C:Superfamily: C1r/C1s repeat homology; EGF homology
 F:1-19/Domains: signal sequence #status predicted <Sig>
 F:10-1064/Product: fibropellin I #status predicted <FIB>
 F:23-54/Domains: EGF homology <EG01>
 F:57-175/Domains: C1r/C1s repeat homology <C1R>
 F:180-211/Domains: EGF homology <EG02>
 F:218-249/Domains: EGF homology <EG03>
 F:256-287/Domains: EGF homology <EG04>
 F:294-325/Domains: EGF homology <EG05>
 F:332-363/Domains: EGF homology <EG06>
 F:370-401/Domains: EGF homology <EG07>
 F:408-439/Domains: EGF homology <EG08>
 F:446-477/Domains: EGF homology <EG09>
 F:484-515/Domains: EGF homology <EG10>
 F:522-553/Domains: EGF homology <EG11>
 F:560-591/Domains: EGF homology <EG12>
 F:598-629/Domains: EGF homology <EG13>
 F:636-667/Domains: EGF homology <EG14>
 F:674-705/Domains: EGF homology <EG15>
 F:712-743/Domains: EGF homology <EG16>
 F:750-781/Domains: EGF homology <EG17>
 F:788-819/Domains: EGF homology <EG18>
 F:826-857/Domains: EGF homology <EG19>
 F:864-895/Domains: EGF homology <EG20>
 F:902-933/Domains: EGF homology <EG21>
 F:936-1064/Region: avidin-like
 F:33-34,28-43,45-54,62-88,180-191,185-200,202-211,218-229,223-238,240-249,256-267,261-275,453-466,468-477,484-495/Duplicate bonds: #status predicted
 57,489-504,506-515,522-533,527-542,544-553,560-571,565-580,582-591,598-609,603-618,620-624,629-640,642-653,655-666,668-679,681-692,694-705,707-718,720-731,733-744,746-757,759-770,772-783,785-796,798-809,811-822,824-835,837-848,850-861,863-874,876-887,889-900,902-913,907-922,924-933/Duplicate bonds: #status predicted
 08,810-819,826-837,831-846,848-857,864-875,869-884,886-895,902-913,907-922,924-933/Duplicate bonds: #status predicted

Query Match 14.28; Score 954.5; DB 2; Length 1064;
 Best Local Similarity 28.0%; Pred. No. 1.6e-43;
 Matches 290; Conservative 93; Mismatches 305; Indels 347; Gaps 69;

QY 30 DPVWGHMSYSTVQESYHPPDQIYVSTCLINNFKTRRVRVSTRYRGEKTMV 89
 DB 181 DPNLCONG-----AACTDLVNDYACT----- 201
 QY 90 RKSQCPGFESGEMC---VPHCA-DKCVHRCIAPN---TCQCEPGMGTCSSACDG 141
 DB 202 -----CPGPF--TGRCNEIDIDECASDPQNGACVGVNVCNCPVGFSGDECCENIN- 254
 QY 142 DHMGPHCTSGCKNGALCNPIRGA---CHCAAGFRGWACE--DRCCEGTGNDCHQR 194
 DB 255 -----ECAS--SPCLNGIGIC--VDGVNMFECTCLAGFTGVACEVINDECCAS----- 296
 QY 195 COQONATC-DHYTG--ECRPPRYTGAFCEDDLPPGKHGQCGORCPQNGVCHVYTG 252
 DB 195 COQONATC-DHYTG--ECRPPRYTGAFCEDDLPPGKHGQCGORCPQNGVCHVYTG 252

DB 297 APCONGICIDINGTYCSCPLGSGDNCN-----NDECCS-IPCLNGTCVDLVNA 349
 QY 253 --CSPSGMWGVTCGGPCCPGRGFRGNKNCOSQOCCHNGTC-DATG-QCHSPGYTGERQ 308
 DB 350 YMCVCAPGWTGPTCADNIDE-----CA-SAPONGICIDYNGVWCDCQPPYTGTHCE 402
 QY 309 ---DECPVGTGVYCAETCCOVNGKCYH-VSG-ACLCENAGFAGERCARLCEPLGYIK 363
 DB 403 TDIDEC-----ARPPCQNGGDCVDGVNVCICAGFGLNCE----- 440
 QY 364 CDRRCCHLENTHSC--HPMSGECACKPRHSGLYCNETGSPGYGACQ--QICS---C 415
 DB 441 -----NNIDECASRPQONGAVCYDGVNGFVC--TCSAGTGVLCEDINECASMPC 469
 QY 416 QNGADC-DSVTGK-CHCAPGFKGIDSTPCPLTYGINGSRCGKNDAVCS-PVGG-SC 471
 DB 490 LINGVCTDLVNGYTCACAGFEGTNCETDID-----ECAS-PPCQNGARCTQVNVGYC 542
 QY 472 TCRAHGHVDC-----SIRCPG-----TWFGCNL-----TCQ----- 500
 DB 543 TCVPGYTVLCEDINECASFPCLNGTCDNDVNGVCAQDTSVSTETDRDECASAP 602
 QY 501 CLNGGAC-NTLDG-TCTCAGMGRKCEL-----PCQDGYGLNCAERCCSHADGC 550
 DB 603 CLNGGACMDVNVNFCVCTCLPGWEGTNCETIDECASSPCNG--GL-CVDQVN-STV--- 655
 QY 551 HPTTGRCRLPGWSGVHCDSCVCAEGRMGPNCSLPCYCKNGASG--SPDDGICECAGFRG 608
 DB 656 -----CFCLPFGTGHGCEIDECCASSP-----CLNGGQCLDNDVSDYECVCAAGYTA 702
 QY 609 TTCQ-----RIGSPFGYHRCSTQCPQCHSSGPC----- 638
 DB 703 VRQOINDECASAPCONGVGVGVNGVYCNCAFGYTGDCETIDEIC--ASMPCLNGGA 760
 QY 639 --HHNG-LGDCPLGFGALC-----NEVPSGFRGNCAIGCTNNGTCNPIDRS 667
 DB 761 CIEMVNGYTCQVAGTGTGCEIDIDECCASAPONG-----GVCTDITNGTY----- 807
 QY 688 CQCPGWIGSDCSQPCPPAHMGPNCTHTCNHGAFCASAYDG--ECKCTPGMTGLYCP 743
 DB 808 CACVPFGTSGNCTNIDECCASDP-----CLNGGIC--VDGVNGFVFCQCPHPSGYTCE 858
 QY 744 QRCPLGFGKDCALICQCONGALCDBHISG--CTCRGTGFRGCE--QKCPGTYGYC 798
 DB 859 -----ISLDACRSMPCQNGATCVVGVADYCECPGVAGNCEIDINECAS----- 904
 QY 799 RQICDCLNNSGTC-DHITG--TCYCSPGMKGARCOAGYIIVGN-----NSLSRTSTA 848
 DB 905 ---LPCQNGSLCIDGAGTGCQRLGTYGVNCEVGFCDLGMVMECNDQVITRTSTG 961
 QY 849 LPADSYOIGAIAGIILVVLFLALFTIYRHKOKGESSMPAYVYTPAMRYVNDYTI 908
 DB 962 M-----MLGDYMYNBRALGYAPVYVGVASN-----NNDPFS 995
 QY 909 SG-TLPHSNGGNANS 922
 DB 996 FGFTVVRDNGQSTTS 1010

RESULT 15
 T09059
 notch4 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jan-2000
 C:Accession: T09059
 R:Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Loretz, C.;
 submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: Z16543
 A:Accession: T09059
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1964 <ROW>

A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564947

C:Genetics:

A:Gene: notch4

A:Map position: 17

A:Intons: 22/1; 148/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67

1679/3; 1729/1; 1761/3

C:Superfamily: unassigned EGF-related proteins; EGF homology

C:Keywords: receptor; signal transduction

F:514-545/Domain: EGF homology <EGF>

Query Match 14.1%; Score 952.5; DB 2; Length 1964;

Best Local Similarity 26.5%; Pred. No. 3,6e-43;

Matches 298; Conservative 68; Mismatches 315; Indels 445; Gaps 65;

```

QY 95 CP-GFESEEMCVPHCADCV-----HRCIAPNT-----CQCEPQWGTGNCSSACDGH 143
DB 102 CDSGF--TGDRQTHLEELCELPSPFCSSNGCHCYVQASGRQPCSEPEMTQEQQLR----- 154
QY 144 WSPHCTSRQCKNGALCNFTYG--ACHCAAGFRGNCE-----DRCEQGTGNDG 191
DB 155 --DFCSAN--PCANGGCVLATYPOIQRCRPFEGHTECENDINECFLEPQPCFQGT---SC 208
QY 192 HQ-----RCQ-----CONGATCD---HVTGE--CRCPGTYGA 219
DB 209 HNTLASVQCLCPVGEQPCQKLNKACAPGSCINGGTCQLYPEGHSTFHLCLCPGFTGL 268
QY 220 FCE-----DLCPGKHG-----PQCEORCP--CONGV 245
DB 269 DEEMNDDCVRHOCNGATCLDGLDITYTCLCPRTWKGMDCSBDIDCEARPPRCRNGT 328
QY 246 CHHVTG--ECSPGWMGTVCQGP-----CPEGRFGKNC 278
DB 329 CNTAGSFHCVCYSGWGGAGCEENLDCAATCABGSTCIDRVGSTRSCLCPRGTGLCH 388
QY 279 QE--C--OCHNGTC--DAATGQ--CHCSPGYTGERCQ--DECPVGTGYVCAETCQC 326
DB 389 LEDMCISOPCHVNAOCSTNPLTGSTLICQPGYSGSTCHQDLDECQMAQGG-----PSPC 443
QY 327 VNGGRCYHSGA--CLCEAGFAGERCEAR-----LCPEGIX 360
DB 444 EHGGSCTINPGSPNCLCLPGYTSRCEADHNECLSQPCHPGSTCLDLATFHLCLCPPGLE 503
QY 361 GIKCD---KRC---PCHLEHTSHCHPMG--ECACKPGMSGLYCNE----- 398
DB 504 GRUCEVEVNECTSNPC--LNQAAHDLNGFOCLCLPFTGARGCEKMDCECSTPCANG 561
QY 399 -----TCSPEFYGEACQOICS-----CONGADCDVYTK--CTCAPGFKGIDC 439
DB 562 RCRDQPAFYCECLPGEFGEHCEKEVDECLSDPCPYGASCLDLPGAFFCLCRPGFTGLC 621
QY 440 STP-----CPLGTYGI-----NCSSRCG----- 457
DB 622 EYPLCTPNMCPQPGQCGQGEHRAPECLCPDGSFPCVPAEDNCPCHHGHGCRSLCVCDEGWT 681
QY 458 -----CKNDVCSPYDG--SCTCKAGMHGVDG--IRCPSGTW----- 491
DB 682 GPBCETELGGCISTPCRAHGTCHPQPSGYNCTCPAGYMKLTCSBEVYACHSGPCLNGSSC 741
QY 492 -----GFGCN-----LTQCLNGACNTLDGT--CTCAPNGRGEKE 526
DB 742 SIRPEGYCTCLPSHTGRHCQTAVDHCVASCLNGTCVKNKPTFFCLCATGFQGLHCE 801
QY 527 -----LPCQGTGILNC-----AERCDSSHADGCHPT--- 553
DB 802 KTNPSCADSPCKNKATCQDTPRGARCLCSPGYTSSCQTLIDLCAKPCPHARCLQSGP 861
QY 554 TGHCRCLPQMGVHCD--SVCAEGRWGPNCSLPCYCKNGASCSPDG--ICECAPFRG 608
DB 862 SFQCLCLQGWMTALCDPFLPSCKAKAMSGIETISGLCQNGGLCT-DTGSSTFCRCRPGFG 920
QY 609 TTCQRICSPGYGHRCSQTCPOCVHSSGPCHHITGLCDCLPFTGALCNEVCPSGRFGKN 668
DB 921 KLCQDNVNP-----C--EPNCHHGS--TCVPOPSGYVCO--CAPGYEGGN 960

```

```

QY 669 CAGI---C---TCTNNGTC--NPIDRSCQCYPGMTGSDC-----SQPCPPAHMGRNC 712
DB 961 CSKVLDAQCQSPCHNHGTCTSRPGRGFCACPPGFVGLRCEGVNDECLDRPCHPS----- 1014
QY 713 IHTCNCHNGAFCSAYDDECKCTPGMTGLYC-----TQ 744
DB 1015 -GTACH--SLNAF--TYQCLPGHTGQRCVEYEMDLQSQSPCSNGSGCEITTPPGFTC 1069
QY 745 RCPLEFYGRKC---ALIC---OCNGADC-----DHISQCTCRTGFMGRHC--EOKCPSG 792
DB 1070 HCPKGFEPGTCSHKALSCGIIHHCHNGGLCLPSKPCSPPLCACLSGFGSPDCLTPAPP- 1128
QY 793 TYGCGRCQICDCLNSTCHITG-----TYCSPGMRKARCDQAG 832
DB 1129 ---GCGPPSPCLHNGCTETPEGLNPGFQCTCPDPSPPRCQRPQ 1170

```

Search completed: May 9, 2003, 11:58:26
Job time : 64.0243 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:54:50 ; Search time 122.851 Seconds

(without alignments)
853.959 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744

Sequence: 1 MVISINSLSTFICLLCHWT.....SSPKQEDSGSSSSSSSS 1140

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6744	100.0	1140	US-10-092-390-2	Sequence 2, Appl1
2	3601	53.4	586	US-10-092-390-4	Sequence 4, Appl1
3	2506.5	37.2	1050	US-09-796-753-114	Sequence 114, App
4	1909	28.3	636	US-09-796-753-100	Sequence 100, App
5	1909	28.3	636	US-09-796-753-124	Sequence 124, App
6	1192	17.7	474	US-10-084-994-11	Sequence 11, Appl1
7	1034.5	15.3	2444	US-09-944-849-2	Sequence 2, Appl1
8	1011	15.0	241	US-10-084-994-8	Sequence 8, Appl1
9	961.5	14.3	497	US-09-796-753-116	Sequence 116, App
10	870.5	12.9	830	US-09-870-759-134	Sequence 134, App
11	870.5	12.9	830	US-09-870-759-140	Sequence 140, App
12	870.5	12.9	830	US-09-842-758-57	Sequence 57, Appl1
13	815	12.1	639	US-09-900-449A-4	Sequence 4, Appl1
14	808	12.0	865	US-09-842-758-20	Sequence 20, Appl1
15	795.5	11.8	601	US-09-900-449A-5	Sequence 5, Appl1
16	790	11.7	1055	US-10-219-248-2	Sequence 2, Appl1
17	790	11.7	1055	US-10-219-247-2	Sequence 2, Appl1
18	790	11.7	1055	US-09-855-722-2	Sequence 2, Appl1
19	789	11.7	1212	US-10-219-248-3	Sequence 3, Appl1

20	789	11.7	1212	9	US-10-219-247-3	Sequence 3, Appl1
21	789	11.7	1212	10	US-09-855-722-3	Sequence 3, Appl1
22	789	11.7	1238	10	US-09-944-849-4	Sequence 4, Appl1
23	783.5	11.6	934	9	US-09-842-758-18	Sequence 18, Appl1
24	779	11.6	296	9	US-09-866-050A-458	Sequence 458, App
25	779	11.6	299	9	US-09-866-050A-192	Sequence 192, App
26	779	11.6	1238	9	US-09-866-050A-332	Sequence 332, App
27	779	11.6	1238	9	US-10-219-248-5	Sequence 5, Appl1
28	779	11.6	1238	9	US-10-219-247-5	Sequence 5, Appl1
29	779	11.6	1238	10	US-09-855-722-5	Sequence 5, Appl1
30	769	11.4	2201	12	US-10-100-912-2	Sequence 2, Appl1
31	768.5	11.4	572	9	US-09-900-449A-7	Sequence 7, Appl1
32	766	11.4	2167	10	US-09-778-927A-61	Sequence 61, Appl1
33	757	11.2	1036	10	US-09-995-593A-6	Sequence 6, Appl1
34	757	11.2	1187	10	US-09-995-593A-7	Sequence 7, Appl1
35	757	11.2	1208	9	US-10-213-329-1	Sequence 1, Appl1
36	757	11.2	1218	10	US-09-995-593A-11	Sequence 11, Appl1
37	757	11.2	1218	10	US-09-944-849-3	Sequence 3, Appl1
38	747	11.1	1218	9	US-10-219-248-7	Sequence 7, Appl1
39	747	11.1	1218	9	US-10-219-247-7	Sequence 7, Appl1
40	747	11.1	1218	10	US-09-855-722-7	Sequence 7, Appl1
41	739.5	11.0	566	9	US-09-900-449A-6	Sequence 6, Appl1
42	735.5	10.9	2743	9	US-10-037-182-6	Sequence 36, Appl1
43	735.5	10.9	3695	9	US-10-037-182-2	Sequence 2, Appl1
44	721.5	10.7	310	9	US-10-084-994-12	Sequence 12, Appl1
45	717.5	10.6	1801	10	US-09-938-275-8	Sequence 8, Appl1

ALIGNMENTS

```
RESULT 1
US-10-092-390-2
: Sequence 2, Application US/10092390
: Publication No. US20030013865A1
: GENERAL INFORMATION:
: APPLICANT: Yu, Xuanchuan
: TITLE OR INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucle
: FILE REFERENCE: LEX-0317-USA
: CURRENT APPLICATION NUMBER: US/10/092,390
: PRIOR FILING DATE: 2002-03-06
: PRIOR APPLICATION NUMBER: US 60/275,013
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 1140
: TYPE: PRT
: ORGANISM: homo sapiens
US-10-092-390-2

Query Match      100.0%; Score 6744; DB 9; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVISINSLSTFICLLCHWIGTASPLNLEDPVNCCHMSYSTVQESYPRHPDQIYRSC 60
DB 1 MVISINSLSTFICLLCHWIGTASPLNLEDPVNCCHMSYSTVQESYPRHPDQIYRSC 60
QY 61 TDILWFMCCTHRRYSRYRATYRGEKTYRRKSQCCPGYESEGMKVPHPCADKCVGRGIA 120
DB 61 TDILWFMCCTHRRYSRYRATYRGEKTYRRKSQCCPGYESEGMKVPHPCADKCVGRGIA 120
QY 121 PNTCCCEPGEWGTNCSSACDDDHMGPHCTSRQCCKNGALCNPTGACHCAAGFRGWRCD 180
DB 121 PNTCCCEPGEWGTNCSSACDDDHMGPHCTSRQCCKNGALCNPTGACHCAAGFRGWRCD 180
QY 181 RCEGTATNDCOROCNGATCHVTGECRCPPRYTAFEDLCPPKHGPOCORPC 240
DB 181 RCEGTATNDCOROCNGATCHVTGECRCPPRYTAFEDLCPPKHGPOCORPC 240
QY 241 ONGGVCHVTGECRCPSGGMGTVCGPCPEGRFGKNCSEQCHNGTCTDAATGCHCSP 300
```

```

|||||
Db 241 QNNGVCHHTYGECSPPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCCAAAGQCHCSP 300
OY 301 GTTGERCODECPYGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
Db 301 GTTGERCODECPYGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
OY 361 GIKCDRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGYEACQOLSCONGAD 420
Db 361 GIKCDRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGYEACQOLSCONGAD 420
OY 421 CSVTGKCTCABGFKIDISTPCLGTGTCGSCRCCKNDAYCSYVDSCTCAKMHGY 480
Db 421 CSVTGKCTCABGFKIDISTPCLGTGTCGSCRCCKNDAYCSYVDSCTCAKMHGY 480
OY 481 DCSIRPSTGTWGRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPCODGTGGLNCAE 540
Db 481 DCSIRPSTGTWGRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPCODGTGGLNCAE 540
OY 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPYCKNGASCSPDDGIC 600
Db 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPYCKNGASCSPDDGIC 600
OY 601 ECAPGRGTTCCORICSPGYTHGRCSTQCPCVHSSGPRCHHTITGLCCLPEFTGALCNEVC 660
Db 601 ECAPGRGTTCCORICSPGYTHGRCSTQCPCVHSSGPRCHHTITGLCCLPEFTGALCNEVC 660
OY 661 PSGRFKNAGICTCTNNGCNPIIDRSQCPCYPMISDSCQPCPAHMGNCIHTCNCH 720
Db 661 PSGRFKNAGICTCTNNGCNPIIDRSQCPCYPMISDSCQPCPAHMGNCIHTCNCH 720
OY 721 GAFCSAYDECKCTPGMTGLYCTQRCPLGFYGRKDCALICQCONGADCDHISGQCTCRGTG 780
Db 721 GAFCSAYDECKCTPGMTGLYCTQRCPLGFYGRKDCALICQCONGADCDHISGQCTCRGTG 780
OY 781 MGRHCKOKPSPGYTGCGQICDCLNNSCTDHTTCTCSPGKRGARCOAGIYIGNLN 840
Db 781 MGRHCKOKPSPGYTGCGQICDCLNNSCTDHTTCTCSPGKRGARCOAGIYIGNLN 840
OY 841 SLSTRSTALPADSYOIGALAGIILVLFLLALFIYRHKOKGESSMPATYTPAMR 900
Db 841 SLSTRSTALPADSYOIGALAGIILVLFLLALFIYRHKOKGESSMPATYTPAMR 900
OY 901 VYNADYITIGTLPHSNGNANSHYFTNPSTHTLTQCATSPHVNRRDMATYTSKNNQLEY 960
Db 901 VYNADYITIGTLPHSNGNANSHYFTNPSTHTLTQCATSPHVNRRDMATYTSKNNQLEY 960
OY 961 NLKNVAPGRGPGVDCGTGLPADMKHGYLNEIGAFGLDRSYVSKSLKXDKGSEYNSN 1020
Db 961 NLKNVAPGRGPGVDCGTGLPADMKHGYLNEIGAFGLDRSYVSKSLKXDKGSEYNSN 1020
OY 1021 CSLSSENYPATIKDPVILPKSSECGYEMKSPARRDSPYAEINNSTANRWYEVEPT 1080
Db 1021 CSLSSENYPATIKDPVILPKSSECGYEMKSPARRDSPYAEINNSTANRWYEVEPT 1080
OY 1081 VSVYGVCFVSNRGLSODPDLPRNHSHPCHYDILLPYRDSSSSPKQDSGSSSSSSSE 1140
Db 1081 VSVYGVCFVSNRGLSODPDLPRNHSHPCHYDILLPYRDSSSSPKQDSGSSSSSSSE 1140

RESULT 2
US-10-092-390-4
; Sequence 4, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucleotide
; FILE REFERENCE: LEX-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12

```

```

; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4
; LENGTH: 586
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-4

Query Match          53.4%; Score 3601; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 1e-168;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVLINSLCSFLICLLCHNIGTASPLNLEDPNVCMSHESYSTVQBSYRHPEDQIYITSC 60
Db 1 MVLINSLCSFLICLLCHNIGTASPLNLEDPNVCMSHESYSTVQBSYRHPEDQIYITSC 60
OY 61 TDLINFKCTRRHVSRTAYRHEKMYRRKSQCCGFTESGMCYPRHCAKCVHRCIA 120
Db 61 TDLINFKCTRRHVSRTAYRHEKMYRRKSQCCGFTESGMCYPRHCAKCVHRCIA 120
OY 121 PNTCOCEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
Db 121 PNTCOCEPMGNTGNCSSACDGDHMGPHCTSRCCCKNGALCNPTTGACHCAAGFRGRCD 180
OY 181 RCEQGTGYNDCHQRCQCONGATCDHYTGECRCRPPGYTGAFCEBDLCPPGRHGPCEQRCP 240
Db 181 RCEQGTGYNDCHQRCQCONGATCDHYTGECRCRPPGYTGAFCEBDLCPPGRHGPCEQRCP 240
OY 241 QNNGVCHHTYGECSPPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCCAAAGQCHCSP 300
Db 241 QNNGVCHHTYGECSPPSGMMGTVCQPCPEGRGKNCSECCQCHNGTCCAAAGQCHCSP 300
OY 301 GTTGERCODECPYGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
Db 301 GTTGERCODECPYGTGVLCAETCCQCVNGKCYHVSAGACCEAGFAGERCEATLCEGLY 360
OY 361 GIKCDRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGYEACQOLSCONGAD 420
Db 361 GIKCDRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGYEACQOLSCONGAD 420
OY 421 CSVTGKCTCABGFKIDISTPCLGTGTCGSCRCCKNDAYCSYVDSCTCAKMHGY 480
Db 421 CSVTGKCTCABGFKIDISTPCLGTGTCGSCRCCKNDAYCSYVDSCTCAKMHGY 480
OY 481 DCSIRPSTGTWGRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPCODGTGGLNCAE 540
Db 481 DCSIRPSTGTWGRGCMILTCQCLNGGACNTLDGTCTCAPGRGKCELPCODGTGGLNCAE 540
OY 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPYCKNGASCSPDDGIC 600
Db 541 RCDCHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPYCKNGASCSPDDGIC 600

RESULT 3
US-09-796-753-114
; Sequence 114, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26

```

```

; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 114
; LENGTH: 1050
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-796-753-114

```

```

Query Match      37.2%; Score 2506.5; DB 9; Length 1050;
Best Local Similarity 40.5%; Pred. No. 3 9e-115;
Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

```

```

QY 14 LLLCHWIGTASPLNLEDPNVCSHWESYSTVQESYPPHFDQIYTTCTDILM-----66
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 9 LLLAVGRLAGTLNPSDPNCSFEWSESTTTTKESHSPFLSPERCE--RPMGPHATCP 66
QY 67 -----FKCRHRYSVR-----TAY 80
DB 67 SPOTORKLASRDSFCVAVCVAGAGVQWRDRSALOPOTGNALSMRPPRYLSGASLSAPGH 126
QY 81 RHGEKTYARRKSQCPPEYSGEKCVPYHCAKDCVHGRCIAPNTCCGEPGNGTNCSSA--138
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 127 TVVVKTDHRRKRLQCCCHGYFSRGFCVPLCAQECVHGRCVAPNCCQCVPGNRGDDCSSAPN 186
QY 139 -----CDGDHNGPHCTSRCKCKNGLCNPLTGACHCAAGFGNWCEDRCDEGTGNDCHOR 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 187 CLAPCTPGYGPACQFCRCQ--HGAPCDPQTGACFCFAERTRGPGCDVSSCGT-----237
QY 195 CQCNQATCTDHTVGECCRPCTGACFCEDLCPRGKKGPOCEQKPCQNGCVCHHVTGEC 254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 238 -----SGFTC-----PSTH-----PCQNGGVOTPOGSSCS 262

```

```

QY 255 CPSSMMGTVCOPCPREGRFKNCSOECQCHNGGTCDAATGCHSCPSGTGERCODECPVG 314
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 263 CPPEGMGTICSLPCRESEFHRPNCSQECRCNNGICLDFEQCKCARPETYTDKRECEPVG 322
QY 315 TYGLCAETCCQVNGKCVHVSAGACLEAGFAGEBCEARLCPRGLYIKDKRCPCHLEN 374
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 323 RFGQDCAETCDCAADAPACFPANAGACLEHGFETGDRCTDRICLPDGFVYLSQAOACTOREH 382
QY 375 THSCHPMSGECACKPGMSGLYCNETCSPGFYGEACQOICSCQNGACDSDYTKCTCAPGF 434
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 383 SLSCHPMNGECSCLPGVAGLHCNESCPODRHNGECCEHCLCLNGVCOAATSGLOCAAPGY 442
QY 435 KGJDCSPRCPLGTGYINCSSRCGCKNDAYCSPVDSCTCKAGMHGVNCSIRCPSTGTAFGC 494
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 443 TGRHCSLCPDPTTYGVNCSKSCENALIASPDIEGVCLEGWGRGNCSPCPGTWFGS 502
QY 495 CNLTQCLNGACNTLDGTCTCAPWNGECCCELLPCODGTGYLNCACRDCSDHADGCHPTT 554
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 503 CNMSQCAHBAVCSPOGGACTCTPRMNGAHQCLPCPKGQFGECCASRCDCHSDGCDPVH 562
QY 555 GHRCRLPMSGVCNCDSCAGRWGPNCSLPCYCKNAGNSCSPDDGICCCAGFRGTTCORT 614
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 563 GRCQCCQAGMAGARCHLSCPRGLMGVNCNNTCTCKNGGTCLPENGNCVCACAFGRGSCORS 622
QY 615 CSPGFYGHRCSTQPCQCVHSSGCHHTGLCDLPGFTGALCNEVCPSGRFKNACICT 674
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 CQFGRYKRC-----CVP-----CK 636
QY 675 CTNNGTCNPIDRSQCCYPMIGSDCSQPCPPAHMPCNICTCNCNAGFCSAYDGECKCT 734
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 637 CANHSFCHPNSNGICYCLAGWTGPDSCQPCPGHNGENCACTCCCHHGCTCHPDGSCICP 656
QY 735 PGMWTGLCTQRCPLGFGKCALICQONAGADCDHIGGCTCKRFGFRCEQKCPSGTY 794
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 697 LGMTGHLCEGCRPGTGANCSPQCGPGEKC-----HPE-----732
QY 795 GYGCQRQICDCLNNTCDHNGTCYCSPGMKGARCDAQAGVILGNLSSTSTALPAD--852
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 733 -----TGACVCPRHSGAPCR-----IG-----IQEPHYMPTTPV 763
QY 853 SY-QIGAIAGIILVLYLFLALFIYLRHKOKRESSMPAVYTTTPAMRVVNDYTTISGT 911
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 764 AYSISGAVIGIAGVLSVVALVLFYGRHMOKKEHNNHVAVYSSG--RLDSEYVAPDV 822
QY 912 LPHNSGNANSHYFTNYSYTLFOCATSPHVNRRDRTVTKSKNNOFLVULK--VNGCR 970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 823 PP-----SYSHYTSNYSHTLSCSPNPPPK-----VPGPLFASLQNERPG--866
QY 971 GPVG-DCTGLTPADMKH-----GYLNLGAFGLDRSGYMGKSL-----KDLGKNSEY 1016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 867 GAGGHDNHTLPLADMKHRRRPPGPIADR--GSSRLDRYSYSYSGNPGPFYDKLISREEL 925
QY 1017 NSSNCSLSSENFPATIKDPPVILPKSSECGYEMKSPARDDSPYAEINNSTANRVYE 1076
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 926 GASVAST--SEENPYATIRDLPSLPGCRRESSYEMKPGSAPROPQFWDSDORRR---981
QY 1077 VEPTVSVQGVFSNNGUL-----SQDP-----VDLPKNSHIPPCHYDLPLVRDS 1119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 982 -QPOPORDSGTQPSPLIHDRDSVGSQPLPGLPGRGHYDPSKNSHIPGHYDLPLVRHP 1040
QY 1120 SSSP-KOED 1127
DB 1041 PSPPLRQD 1049

```

```

RESULT 4
US-09-796-753-100
; Sequence 100, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999

```

;; CURRENT APPLICATION NUMBER: US/09/796,753
;; CURRENT FILING DATE: 2001-03-01
;; PRIOR APPLICATION NUMBER: 09/183,175
;; PRIOR FILING DATE: 1998-10-30
;; PRIOR APPLICATION NUMBER: 09/223,094
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/223,546
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/224,246
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/259,388
;; PRIOR FILING DATE: 1999-02-26
;; PRIOR APPLICATION NUMBER: 60/122,458
;; PRIOR FILING DATE: 1999-03-01
;; PRIOR APPLICATION NUMBER: 09/312,359
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 09/336,536
;; PRIOR FILING DATE: 1999-06-18
;; PRIOR APPLICATION NUMBER: 09/342,687
;; PRIOR FILING DATE: 1999-06-29
;; PRIOR APPLICATION NUMBER: 09/345,464
;; PRIOR FILING DATE: 1999-06-30
;; PRIOR APPLICATION NUMBER: 09/365,164
;; PRIOR FILING DATE: 1999-07-30
;; PRIOR APPLICATION NUMBER: 09/399,723
;; PRIOR FILING DATE: 1999-09-20
;; PRIOR APPLICATION NUMBER: 09/409,634
;; PRIOR FILING DATE: 1999-09-30
;; PRIOR APPLICATION NUMBER: 09/471,179
;; PRIOR FILING DATE: 1999-12-23
;; PRIOR APPLICATION NUMBER: 09/474,071
;; PRIOR FILING DATE: 1999-12-29
;; PRIOR APPLICATION NUMBER: 09/474,072
;; PRIOR FILING DATE: 1999-12-29
;; PRIOR APPLICATION NUMBER: 09/514,010
;; PRIOR FILING DATE: 2000-02-25
;; PRIOR APPLICATION NUMBER: 09/516,745
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: 09/572,002
;; PRIOR FILING DATE: 2000-05-14
;; PRIOR APPLICATION NUMBER: 09/597,993
;; PRIOR FILING DATE: 2000-06-19
;; PRIOR APPLICATION NUMBER: 09/599,596
;; PRIOR FILING DATE: 2000-06-22
;; PRIOR APPLICATION NUMBER: 09/630,334
;; PRIOR FILING DATE: 2000-07-31
;; PRIOR APPLICATION NUMBER: 09/606,565
;; PRIOR FILING DATE: 2000-06-29
;; PRIOR APPLICATION NUMBER: 09/606,317
;; PRIOR FILING DATE: 2000-06-29
;; PRIOR APPLICATION NUMBER: 09/665,666
;; PRIOR FILING DATE: 2000-09-20
;; PRIOR APPLICATION NUMBER: 09/677,751
;; PRIOR FILING DATE: 2000-09-30
;; NUMBER OF SEQ ID NOS: 162
;; SEQ ID NO 100
;; LENGTH: 636
;; TYPE: PRT
;; ORGANISM: Rattus sp.
US-09-796-753-100

Query Match 28.3%; Score 1909; DB 9; Length 636;
Best Local Similarity 45.1%; Pred. No. 3.5e-86;
Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;

QY 260 NGTVCGPCGPRGKNGKNSOGCHNGTCAATGCGCHSGVNGERCODECPVTGYVL 319
DB 1 MGVCISLPCPGRFHPNCTGCRKCHNGSLCDRTGQCHCAPGYIGDRCRECPVGRFGOD 60

QY 320 CAETCCVNGGKCVHVSACLCCEAGFAGERCERARLCPEGLYGIKCDKRCPCHELTSHSCH 379
DB 61 CAETCCDCAFGARCRFPANACLCCEHGTGDRCTERLCPPDGRYGLSQDQCTCTCPHSLSCH 120

QY 380 PMSGECACKPGMSGLYCNETCSPGFYGEACQICSCONGADSDSYTKCTCAPGFKGIDC 439
DB 121 PMSGECSCQPGMAGLHCHNESCQPDTHGAGCQCHCLHGVGLADBSGLCRAPGYTGPHC 180

QY 440 STPCPLGTGTYGNCSSRCCKNDVAVSPYDGSCTCAGHGVDCSIRCSGTGFGCNCITC 499
DB 181 ANLCPPNTYGINCSHSCCENAIACSPVDTICTCEGNGRNCSPCPGTWGFSCNASC 240

QY 500 OCLNGACNTLDGTCTCAPGRNGEKECELPCCODGTGGLNCAERDCSDSHADGCHPTGHCRC 559
DB 241 QCAHGVCSPTQACTCTPGMRGVACQLPCPRGQDGEBCASVDCDDHSDGDPVHGRC 300

QY 560 LPMGSGVHCDVYACAGRNGPNCSLPCYCRNGASCSPDDICECABGFGTTCQIRICSPGF 619
DB 301 QAGWMTGRCHLPCPEGFAGNCSNACTCKNGGTCTVPENGNCVCARGFGRSPQRRCR 360

QY 620 YGHRCSQTCPOCVHSSGCEHITGLCDLPGFTGALCNEVCPSGRFGKNCAGICTCTNNG 679
DB 361 YGKR-----CVP-----CKCNHS 374

QY 680 TCNPIDRSCOCYPGWIGSDCSQPCPAHMGPNCHTNCNCHNGAFCSAYDGECKTPGMTG 739
DB 375 SCHPSDGTCSCLAGWTGPDSCSPGHMGLKCSQPCQCHGATCHPODGSVCVCPMTG 434

QY 740 LYCTGRCPLEFYKDCALICQONGADCDHISGQCTCRGEMGRHCEKRCPSGTGYGCR 799
DB 435 PNCSEGRSFRMGVNCSQLCCDPPGEMC-----HPE----- 465

QY 800 QICDCLNNSTCDHTGTCTCYCSPGWKGARCDOAGVITVGNLNSLSTSTALPADSYQIGAI 859
DB 466 -----TGACVCPGSHGAKCK-----VSQSEFTIMPTS-PVTHNSLGAIV 504

QY 860 AGIILVVLVFLALFTIYRHKOKRESSMPAVTYTPAMRVNADYITSGTLPHNSNGN 919
DB 505 IGIAVLGTLVALVALFTIGYRMQKGEHEHAAVAVSNG-RIDGSDYVMPDVS----- 557

QY 920 ANSHFTNPSTHTLTQCATSPHVNRRDRTYTKSKNNOLFNLKAVNFGKRCPPVDCGTGT 979
DB 528 SYSHYSPNSYHTLSQCSPPNPPN-----KIPGSQLFVSSQSERPNRHGHDHAT 610

QY 980 LPADMKH 986
DB 611 LPADMKH 617

RESULT 5
US-09-796-753-124
;; Sequence 124, Application US/09796753
;; Publication No. US20030027998A1
;; GENERAL INFORMATION:
;; APPLICANT: McCarthy, Sean A.
;; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
;; FILE REFERENCE: 7853-227-999
;; CURRENT APPLICATION NUMBER: US/09/796,753
;; CURRENT FILING DATE: 2001-03-01
;; PRIOR APPLICATION NUMBER: 09/183,175
;; PRIOR FILING DATE: 1998-10-30
;; PRIOR APPLICATION NUMBER: 09/223,094
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/223,546
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/224,246
;; PRIOR FILING DATE: 1998-12-30
;; PRIOR APPLICATION NUMBER: 09/259,388
;; PRIOR FILING DATE: 1999-02-26
;; PRIOR APPLICATION NUMBER: 60/122,458
;; PRIOR FILING DATE: 1999-03-01
;; PRIOR APPLICATION NUMBER: 09/312,359
;; PRIOR FILING DATE: 1999-05-14
;; PRIOR APPLICATION NUMBER: 09/336,536
;; PRIOR FILING DATE: 1999-06-18
;; PRIOR APPLICATION NUMBER: 09/342,687
;; PRIOR FILING DATE: 1999-06-29

```

; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 124
; LENGTH: 636
; TYPE: PRT
; ORGANISM: Rattus sp.
; US-09-796-753-124

```

```

Query Match      28.3%; Score 1909; DB 9; Length 636;
Best Local Similarity 45.1%; Pred. No. 3,5e-86;
Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;

```

```

OY 260 MGVTCGQPCPEGRFGKNCQEQCHNGTCDATGCHCSFGTGRCDCEPVGYVYL 319
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MGVCISLPCEGFGHPCNQECRCHNGGLCDRTGCHCAPGYIGDRCEECPEVGRFGD 60
OY 320 CAETCCQVNGKCYHNSGACLCCEAGFAGRCERARLCPEGLYGIKDKRRCHELENTSHCH 379
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 CAETCCAGKARCFRANGKACLCCEHGTGDRCTERLCPRDRYGLSCDDPTCDEHLSCH 120
OY 380 PMSGECACKRPMGSLYCNFTCSPGFYGAACQOICSCQNGADSDVTGKCTCARGFKGIDC 439
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 PMHGESCCQPGNAGLCHNCSSCCPODTHGACGQEHCLCLHGVCLADSLRCARAPYGTGPRC 180
OY 440 STPRLGTGINSRSCGCKNDAYCSRYDGSCTCKRAGHGVDSCTSPGSGTGFGCNLTFC 499
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 181 ANLCPRNTYGINSSHCSCENAIACSPVDGTCTCKEGRGNCISVPCPGTWGFSGNASC 240
OY 500 QCLNGAGWTLDGTCTCAGMRGKCELCPCQDGTGTLNCAERDCSHAGCHPTTGHCNC 559
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 241 QCANHEVCSPDTGACTCTGGMGVNQLPCPKGQREBEGSACVDCDDSHSGCPVHGHCHC 300
OY 560 LPMGSGVHCDVCAEBGRWPNCSLPCYCNKNGASCSPDDGICCBAPGFRGTTQRTICSPGF 619
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 301 QAGMMGTCHRLPCPEGFMGANCNACTCKNGGTGCVBENGCYCAPGFRGSPCORPCPRGR 360
OY 620 YGHRCSQTPQCVHSSGPHHTITGLDCLPFTGALCNVVCSSGRFGKCAICTCTNNG 679
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 361 YKRR-----CVP-----CCKNNHS 374

```

```

OY 680 TCNPIDRSCQCYPMGIGSDSCQPCPPAHMGPNCHTNCNHNAGFASAYDGECKTPGWTG 739
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 375 SCHPSDGTSCSLAGWTGPRGSCSCEPFGHMLKCSQPCQCHHRAHTCHPDGSCVCLPRTWG 434
OY 740 LYCTORCPLEFYGCDCALICQONGADCDHISGQCTCRTGFMGRCEOKCPSGTYGYGCR 799
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 435 PNCSECCPSRMRGVNCSQLCCQDDPEMC-----HPE----- 465
OY 800 QICDCLNNSTCHITGTCTCYSGMKGARCDQAGVITVGNLSRSTSLPADSTQIGAI 859
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 466 -----TGACVCPGHSQAMCK-----VGSQSEFTIMPTS-PVITHSLGAV 504
OY 860 AGIILVVLVFLALFIYRHKOKGESSMPPAVTYTPAMRYVNADYITSGTLPHSNGN 919
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 505 IGIIVGLTVLVALFETIRRMOKKEHBLAVASTG-RIDGSDYVWVDP----- 557
OY 920 ANSHFTNPSYHTLQCATSPHYNNDRMTVTKSKNQLFVLKKNVPKRGPPVGDCTGT 979
    ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 558 SYSHYNSNPSTHTLSQCSNPPPP-----KIPGSQLFVSSQASERPRNRHGHDMAT 610
OY 980 LPADMKH 986
DB 611 LPADMKH 617

```

```

RESULT 6
US-10-084-994-11
; Sequence 11, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-Like Polynucleotides, Polypeptides, and Antibodies
; FILE REFERENCE: PTO11PICI
; CURRENT APPLICATION NUMBER: US/10/084,994
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 11
; LENGTH: 474
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-084-994-11

```

```

Query Match      17.7%; Score 1192; DB 9; Length 474;
Best Local Similarity 45.0%; Pred. No. 2,5e-51;
Matches 198; Conservative 43; Mismatches 195; Indels 4; Gaps 4;

```

```

OY 218 GAFCELDLPGRKHGPOCEORCPONGGVCHHTYBSCSPSGMGTVCQPCPEGRFGKNC 277
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 2 GPSCLQACPAAGLYGDNCRHSCLCQNGGTCDPVSGHACPEGNAGLACEKEKCLPRDYRAGC 61
OY 278 SQEQCHNGGTDAATGCHCSPTGTGERCQDECPVGYTGLCAETCCQVNGKCYHNSG 337
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 62 RHSGGCLNGGLDPPHGRILCPAGWTGDKQSPCLRGWGEACARCSCLPMAOHVWG 121
OY 338 ACCEAGFAGRCERARLCPEGLYGIKDKRCHLENTSHSPMSGECACKRGMGSLYCN 397
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 122 ACRCPRGFTGSGCE-QACDPGSGFEGEDCAOMCCQPGENP-ACHPATGTSCCAAGTYGSPSQ 179
OY 398 ETCSBFGYGAACQOICSCQNGADSDVTGKCTCARGFKGIDSTPCPLATYGINSSRGC 457
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 180 QRCPPRGRRGPGEGQLCGCLNGGSCDAATGACRCPGTGTLTDCNLTCPPGGRFGPNCTHYGC 239
OY 458 CKNDAVCSRYDGSCTCKRAGHGVDSCTSPGSGTGFGCNLTGQCLNGACNTLDGTCTCA 517
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 240 CGGAAACDPVYGTCLCPGRAGVRCERGCPOHRRFVGCEHRTCSCHNGGLCHASNGSCSG 299

```


PRIOR FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: PCT/US00/23663
PRIOR FILING DATE: 2000-08-29
PRIOR APPLICATION NUMBER: US 60/151,348
PRIOR FILING DATE: 1999-08-30
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 3.1
SEQ ID NO 8
LENGTH: 241
TYPE: PRT
ORGANISM: Homo sapiens
US-10-084-994-8

Query Match
Best Local Similarity 15.0%; Score 1011; DB 9; Length 241;
Matches 152; Conservative 23; Mismatches 45; Indels 0; Gaps 0;

QY 439 CSTPCLGTGYNCSRCCKNDVAVSPYDGSCTCKAGHGVDCSTIRCPSTGTFEGCNLT 498
DB 5 CAVSCAAGTYGPNCSSTICNNNGTCSPYDGSCTCKEGQGLDCTLPSPSTGWLNCNS 64
QY 499 CCLNGAGANTIDGTCAPGRGECLEPCODGTYGLCAERDCSHADGCHPTTGHR 558
DB 65 CTCAGAACSPIDGSCSTPGMLGDTCELPDGTFFGLNCSEHDCSHADGCDPYTGHC 124
QY 559 CLPGNSGVHCDVCAEGRMGPCSLPCYCKNGASCPDDGICBAGPFGTTCORICSPG 618
DB 125 CLAGWTGICDSTCPGRMRGPNCSVSCSENGSCSPEDGSCBAGPFGPLQRICTPPG 184
QY 619 FYGHRCSOTCPOCVHSSGPCNHTITGLDCLPFTGALCNE 658
DB 185 FYGHCAGCPCLCVHSSRCHHISGICECLPFGSALCNO 224

RESULT 9
US-09-796-753-116
Sequence 116, Application US/09796753
Publication No. US20030027998A1
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
FILE REFERENCE: 7853-227-999
CURRENT APPLICATION NUMBER: US/09/796,753
CURRENT FILING DATE: 2001-03-01
PRIOR APPLICATION NUMBER: 09/183,175
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 09/223,094
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/223,546
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/224,246
PRIOR FILING DATE: 1998-12-30
PRIOR APPLICATION NUMBER: 09/259,388
PRIOR FILING DATE: 1999-02-26
PRIOR APPLICATION NUMBER: 60/122,458
PRIOR FILING DATE: 1999-03-01
PRIOR APPLICATION NUMBER: 09/312,359
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 09/336,536
PRIOR FILING DATE: 1999-06-18
PRIOR APPLICATION NUMBER: 09/342,687
PRIOR FILING DATE: 1999-06-29
PRIOR APPLICATION NUMBER: 09/345,464
PRIOR FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: 09/365,164
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 09/399,723
PRIOR FILING DATE: 1999-09-20
PRIOR APPLICATION NUMBER: 09/409,634
PRIOR FILING DATE: 1999-09-30
PRIOR APPLICATION NUMBER: 09/471,179
PRIOR FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: 09/474,071

PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/474,072
PRIOR FILING DATE: 1999-12-29
PRIOR APPLICATION NUMBER: 09/514,010
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: 09/516,745
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: 09/572,002
PRIOR FILING DATE: 2000-05-14
PRIOR APPLICATION NUMBER: 09/597,993
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 09/599,596
PRIOR FILING DATE: 2000-06-22
PRIOR APPLICATION NUMBER: 09/630,334
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: 09/606,565
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/606,317
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: 09/665,666
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: 09/677,751
PRIOR FILING DATE: 2000-09-30
NUMBER OF SEQ ID NOS: 162
SEQ ID NO 116
LENGTH: 497
TYPE: PRT
ORGANISM: Mus musculus
US-09-796-753-116

Query Match
Best Local Similarity 14.3%; Score 961.5; DB 9; Length 497;
Matches 216; Conservative 53; Mismatches 184; Indels 167; Gaps 15;

QY 545 SHADCHPTTGHRCRLPGNSGVHCDVCAEGRMGPCSLPCYCKNGASCPDDGICECAP 604
DB 2 THASG-DPVHGGCRGADMGTRCHLPCEGPMANCSNITCKNGGTGVSENGHCYCAP 60
QY 605 GFRGTTCCQRTICSPGYGHRCSOTCPOCVHSSGPCNHTITGLDCLPFTGALCNEVCPGR 664
DB 61 GFRGSPCORPCPPGRYGRKCYQ----- 82
QY 665 FCKNAGICTCTNN-GTCNPIRDSOCYPGWIGSDSOPCPAHMGPCNHTCNGNGAF 723
DB 83 -----CKNNHSSCHPDSGTCSCLAGWTGPPDSEACPPGHNGKCSQCLQCHHGT 134
QY 724 CSAVDGECKTPGWTGLYCTQRCPLGFGKDALICQCONGADCHIISQCTCTGFMGR 783
DB 135 CHPDGSCICTPGMTGPNCLDEGCPPMFGVNCSQLQCDLGBMC----- 178
QY 784 HEOQCPSTGYGRCQICDCLANSTCHITTCYCSPPMKARCDQAGVITVGNLSIS 843
DB 179 HPE-----TACVCPFGHSADK-----WGSQEST 205
QY 844 RSTALPADSYQIGAIAGIILVLVFLALFIYRHKOKESMPAYTTPAMRVVN 903
DB 206 IMPTS-PYTHNSLGAIVIGAVLIVALLAIFIGRQMKRKHENHLAVASTG-RDGG 263
QY 904 ADYITSGTLPHSNGNANSHYFTNPSTLTQCATSPHYNNDRMTVTYKSKNQLFVNLK 963
DB 264 SOYVMPDVSP-----SVSHYSSNPSHYTLSSQSPNPPPN-----KVPESQLFVSSQ 310
QY 964 NVNPPKRGVSGCTGTLPADMKHGGVLANELGAFGDRS----- 1002
DB 311 ABERPSRAHGRNHTTLPLDMWHRREPHDRGSHLDRSYSCSYSHRNGBPCHKGPIS 370
QY 1003 --MGSKLDLGRKNSSEYNSNCSLSSSENYATIKDPVYLIPKSSCGYEMKSPARRSD 1060
DB 371 BELGASVMSL-----SSENPYATIRDLPLSGPGRPSGVYEMKGPSPVSP 416
QY 1061 VAEINNSTSANRNYEVER-----TVSVVQGVFSNNGRLSDP-----YDLPRNS 1105
DB 417 RQSLH---LRDQROQLQGRDSDGYEOPSPLSHNEESIGSTPPPLPGLPCHGYDSPKNS 473

```
Qy 1106 HIPCHYDLLPVRDSSSSPKQ 1125
      ||| ||| ||| | | :
Db 474 HIPGHYDLLPVRHPPSPSR 493
```

RESULT 10
100-00-070-750 134

```

: Sequence 134 Application US/09870759
: Patent No. US20020177551A1
:
: GENERAL INFORMATION:
:
: APPLICANT: TENMAN, David S
:
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASES
:
: FILE REFERENCE: 870759
:
: CURRENT APPLICATION NUMBER: US/09/870,759
:
: CURRENT FILING DATE: 2002-01-14
:
: PRIOR APPLICATION NUMBER: US 60/208,128
:
: PRIOR FILING DATE: 2000-05-30
:
: NUMBER OF SEQ ID NOS: 166
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 134
:
: LENGTH: 830
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-870-759-134

```

Query Match	12.9%	Score 870.5	DB 9	Length 830
Best Local Similarity	28.4%	Pred. No. 1.9e-35		
Matches 183; Conservative	45;	Mismatches 154;	Indels 263;	Gaps 22

```

0Y      93  OCCPPEYEGEMC-VPHC--ADKCVHGR-CTAPNTCCCEPBMGTGTTSSACDDHNRPHC 148
0Y      40  OCCAGHRAQDQDCTTPICEBGRDACKDEYCAVGLCRCKPGEFFAHSSNRPGQWYBDC 99
0Y      149  T5RCCKNGALCNPTTGACHCAGAGRWGRNCEDRCEOGTYGNDCHQRCCOQNGATCDVYTG 208
Db      100  RESCECHPHGQCEPATGACCOADRMGRARCEFPACSGPHR-----CDPATG 146
0Y      209  ECRCPRTGTGAFCEBLCFPRKHHGRQCEBRCFQNGVCHHTYGCSCPSGSMGTVCQPC 268
Db      147  VCHCE-----GWSSTCRRP- 162
0Y      269  PEGRGKNCOSQECCHNGGT--CDATGQCHCSPGYTGRCODECPVGTGYVLAEYTCQCV 327
Db      163  -----CQCNTHAAACBQATGACVCKPBMWGR----- 189
0Y      328  NGKCTYHVSGACLEAGFAGERCEARCLPESLYGIKCDKRCPCHLENTHSCHPYSGEAC 387
Db      190  -----CSFRONCH--GSPCEDQSGRCAC 210
0Y      388  KPGMSGLYCNETCSPGFGEACOOICSONGADDSVTGKCTCAPRKIGIDSTPRLGT 447
Db      211  RRGW-----WGRPQDQCEVYKR--CSAASGECTCPGGRGARCELPBPGW 256
0Y      448  YGINSCSRGC--CKNDAYVCSPYDGSCTCKAGHGWYDSCIRPSGTWFGCMILTCQLNGA 506
Db      257  HGVOCAHSCGRCKHNPCSPDTSCE----- 282
0Y      507  CNTLDGTCTCAPGRGKGECELPQDDGTGYGLNCAERK--DCSHADCAHPTTGHK--RCLPRGMS 564
Db      283  -----SCBPPGNGGTCQDQPCLPRTGFGESECBQDQCHHCHGCECEPDTGHCCORDBGW-- 333
0Y      565  GVHCDVYCAEGRMWPNCISLPYCKKNGASCSPPDDQICBARGFPGTTCQIRICSPGFYGHNC 624
Db      334  -----LQPRCEBP-----CPTGTFGEBC 351
0Y      625  SQTCPQCYHSSGPCNHTTGLDCLPFTGALCNEVYCSGFRGKNCAGICTNNGTNCNPI 684
Db      352  GSTCPTCV--OGSCDFTVTDGVCYAGYWGSPSCNAPSAGFHGNNCSVPBCB--EGLCHPY 408
0Y      685  DRSCQCYTP-----WGSQCSQCPRAHNGP 710
Db      409  SGSCQ--PGSGSRPTALVGSVLPFLLLTFGLAC--CACCSNAP 448

```

RESULT 11
US-09-870-759-140
; Sequence 140, Application US/09870759
; Patent No. US20020177551A1
; General Information

Query Match	12.9%	Score 870.5	DB 9	Length 830
Best Local Similarity	28.4%	Pred. No. 1.9e-35		
Matches 183; Conservative	45;	Mismatches 154;	Indels 263;	Gaps 22

QY	93	QCCGFEESGEMC-VPHC--ADKCVHGR-CLAPVTCCEPFGMGSTNCSACDDDHNGPHC	148
Db	40	QCCAGMGNQKODECIIPIICEBGRDACKONDEVCYKBLCSCKGFFGAMHSSSCPRQIWMGRDC	99
QY	149	TSRQCKNGALCNPTGTAGCHACAGFRCMRCEDRCEOGTYGNDCHQRCQCONGATCDHYTG	208
Db	100	RESPCRHPHGOCSEBATGACCOQADRMGAREFFPCACGPHGR-----CDPATG	146
QY	209	ECRCRPGVTAFCEDLCPRGKHGQCSQRCRQCGVCHHYTBECSPSGMMGTVCQGR	268
Db	147	VCHREP-----GMMSTTRRP-1622	
QY	269	PEGRFRKNCSEOCCHNGGT-CDATGQCHCSPTYTERCODEC RVGTGYGLVLAETQCYQ	327
Db	163	-----CQCNMAARCBGATGATACVCKRGMMRR-----189	
QY	328	NGGRCYHSGACILCEAGFAGEBCEARLCRPSBLGIGIKDKRCKRPHLENTHSCHPMSEBAC	387
Db	190	-----CSFRCNCH-----GSPCEODSSRCAC	210
QY	388	KPGWSGLYCNETCSPEGYEGACOOICSCONGADODSVTGKCTCAPRGKIDGTSRPLGT	447
Db	211	RPGH-----WGEBCOQOCCEVAGR-CSASGEBCTPPRFRGRKRLCLPPACG	256
QY	448	YGINCSSRCG-CKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGTFGNCULTQCLNGGA	506
Db	257	HGVQCAHSCGRCKHNECSPDTGSC-----282	
QY	507	CNTLDGCTCAPRGKREKCLPQODGTGYLGNLCERC-DCSHADGCPRTTGHG-RCLEPGNS	564
Db	283	-----SCEPWNMGQCCOQCLPGLTFGESSCDDQCPHCRHGAECBEPDGHCCRCRDPGW-	333
QY	565	GVHCDVSCAEGRMGPNCLSPCYCKNGASCSPPDGIOCECAPGRGTTCCQRTCSGPFYGHC	624
Db	334	-----LGRPCEDP-----CPRTGTGEECC	351
QY	625	SQTCPOCVHSSGPHCHITGLDCLPFTGALCNEVCPSGRFGKNCAGICTCTNNGTGNPI	684
Db	352	GSNTCPCTCV--QGSQDVTGDCVCSAGYMGPSCNASCPAGFHGNNCSVPCECP-EGLCPIV	408
QY	685	DRSCQCCPG-----WIGSSQSPCRPAHNGP	710
Db	409	SSGSCQ--PGSGSDTALINGSLVPLLLLEPLGLAC--CACCCMAP	448

RESULT 12
US-09-842-758-57

```

Publication 57 Application US/09842758
Publication No. US20030083244A1
GENERAL INFORMATION:
APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R.
APPLICANT: Gerlach, Valerie
APPLICANT: Shmukets, Richard A
APPLICANT: Malvankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernhusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kundu
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigarav, Muralidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esna A
APPLICANT: Smithson, Glenda
APPLICANT: Rastelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
TITLE OF INVENTION: No. US20030083244A1 Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 1596-783
CURRENT APPLICATION NUMBER: US/09/842,758
CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,508
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/220,591
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/232,678
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/263,217
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 60/265,160
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 57
LENGTH: 830
TYPE: PRT
ORGANISM: Homo sapiens
US-09-842-758-57
Query Match 12.9%; Score 870.5; DB: 9; Length 830;
Best Local Similarity 28.4%; Pred. No. 1.9e-35;
Matches 183; Conservative 45; Mismatches 154; Indels 263; Gaps 22;
93 QCCGFEYSGMC-VPHC-ADKCVHGR-CIAPNTCOECPGMCNCSACDGHMPPHC 148
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
db 40 QCCAGWRKQDECIITPICEGPDACQKDEVCYKPGILCKRCKPFFGAGHCSRCRPGQIWEFDC 99
149 TSCRQCKNGALCNPIITGACGACAGFGRWCRCEDRCGGTYGNDCHQRCQCGATCDHYTG 208

```

```

Db      100 RESCCHPHGQCEPATGACQCADRMGANCERPCACGPHGR-----CDPATG 146
Oy      209 ECRCPGATGACEDLCPGKHKGRPOCEORCPQNGGVCHNVTGESCPSGMMGTVCQGPC 268
Db      147 VCHCEP-----|||-----GMMSTRRRP- 162
Oy      269 PEGRRGKNCSEQCCHNGST--CDAATGQCHSPGYTGERRCODECPVGTGVLCATETCCCV 327
Db      163 -----CQCNTAARCEQATGACVCYKRGWGGRR----- 189
Oy      328 NGGKCYHVSAGACLCERAGFAGERCEARLCPREGLYGTCIDKRCRCHLENTHSCHMRSGECAC 387
Db      190 -----CSFRCNCH---GSPCEQDSGRGAC 210
Oy      388 KPGMAGLYCNETCSPGFYGEACQOICSCONGADCDVSTKACAPAFKIDCSTPLAT 447
Db      211 RFGM-----WPECCQOCECYRGR-CSAASECICPPFRGARCELPAPAS 256
Oy      448 YGINCSSRGC-CKNDVAVSPVDGSCCTCKAGMHGVDCSIRCPGWTGFGNLATQCCLNGA 506
Db      257 HGVQCAHSGCRKKNRCPDGTGSCF----- 282
Oy      507 CNTLDGTCTCAGMGEKCELPQODGYTGLNCAERC-DCSHADGCHPTGHC-RCLRPGS 564
Db      283 -----SCEFGMWGTQOQDPLRPTGEGSEEQDQPHCRHDEACERDTHGCRDPRGM- 333
Oy      565 GVHCDSVACEGMRGNCNSLPYCKKNGASCSPDGIACAPRFRGTGTCQRICSPGFYHNC 624
Db      334 -----LGRCEDEP-----CPRTGTGEDC 351
Oy      625 SOTCPQCVHSGSPCHHTIGLDCLPRTGALCNEVCPSGRFGKNCAGICTGNTNGICNP1 684
Db      352 GSTCTPLCV--QGSQDTYGDVCYSAGYWGPSCNASCAPAFHGNKSVPECEP-EGLCIRYV 408
Oy      685 DRSCCCTRG-----WIGSDCSQPCPPAHMGP 710
Db      409 SGSCQ--PESGSRDTALIVGSLVPLLLFLFLGLAC--CACCCMAP 448

RESULT 13
US-09-900-449A-4
; Sequence 4, Application US/09900449A
; Publication No. US20030040616a1
; GENERAL INFORMATION:
; APPLICANT: ZHONG, Jenny et al.
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
; FILE REFERENCE: C1001271
; CURRENT APPLICATION NUMBER: US/09/900,449A
; CURRENT FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Strongylocentrotus purpuratus
US-09-900-449A-4

Query Match      12.1%; Score 815; DB 9; Length 639;
Best Local Similarity 30.9%; Pred. No. 7.8e-33;
Matches 236; Conservative 69; Mismatches 242; Indels 216; Gaps 58;

Oy      154 CKAGALCNPTTGA-----CHCAAGFRGMRCE---DRCEQGTGNDCHQRCQONGATC-DH 205
Db      5 CLNGGIC--VDGVNMFECTCLAGFTGVRCENVINDECAS-----APCNGGICIG 52
Oy      206 VTG-ECRCPGATGACEDLCPGKHKGRPOCEORCPQNGGVCHNVTGE--CSCPSSGMMGT 262
Db      53 INGTGTCSPDLGSGNCER-----NDDECSS-IPCLNGTCTVDLVNATMVCAPAGWTGP 105
Oy      263 VCGQCPGPRGRFKNCSQEQCHNGTGC-DAAATG-QCHSPGYTGERCQ---DECPVGTG 317

```

```

Db 106 TCADNIDE-----CA-SAPCGGVGICIDGVNGYMCDCOPGTYGTRHCEIDDEC----- 152
Qy 318 VLCAETCQCGNGKCYH-VSG-ACLEAGFAGEARLCPEGLYIKDKRCPHLENT 375
Db 153 ----APPQNGGDDCVDSVNGYVCAPFEDLNCE-----NNI 187
Qy 376 HSC--HPMSGEACKPGMSGLYCNETCSPEFYGEACQ---QICS---CONGADC-DSVTG 426
Db 188 DECAAPPCNGAVCVGVNGVVC--TCSAGYGVLEETDINECASAPCLNGACADVNA 245
Qy 427 K-CTCAPRGKIDCSFPCGLTYGINSRCCGKNAYVS-PYDG-SCYCKAGMHVDC- 482
Db 246 YICTCAAGEGINTCEDTD-----BCAS-PCQNGATCTDQVNGYVCVPGYTGVLCE 298
Qy 483 -----SIRCPSG-----TWFGCNL-----TCQ-----CLNGGAC-NTLD 511
Db 299 TDINECASFPCLNGGTCNDQVNGYVCADNDSVSTCEIDRDECASAPCLNGACADVNA 358
Qy 512 G-TCTCAPGMRGEKCEL-----PCQDGYGLNCAERCDSSHADGCHPTTGHCRLPG 562
Db 359 GFYCTCLPMEGNTCEINIDECASSPCMNG--GL-CVDQVN-SYV-----CFCLPG 405
Qy 563 WSGVHDSVCAEGRMGNLSLPCYCKNGASCSPDQICCAEGRFRTTCORICSPFYGH 622
Db 406 FTGHCETIDECASSP-----CLNGGQCIDRDVSYEC-----VCAAGYAV 447
Qy 623 RSQSTCPQCVHSSGPCNH-----ITG-LCDCLPFGTALC-NEVCPSGRFGKNCAGIC 673
Db 448 RQGINDEC--ASAPCQNGVCVDVNGYVCNCAPEYTDNCTEL-----DECAAM- 497
Qy 674 TCTNNGTCHPT--DRSCQCYPGHIGSDCSQPCPAHMGPNCHTCHNGHAGCS-AYDGE 730
Db 498 PCLNGGACLEMVNGYVCQVAGTGYVCEIDIDECAASAP-----CONGGVCTDITNGY 550
Qy 731 -CKCPGWTGLVCTORCPGLFYGKDCALICQCONGADC-DHISG-QCTRTGMRHCRQ 787
Db 551 ICACVGFEGNSCENI-----DECA-SDPCLNGGICVDVNGYVCQCPNPSGYCEI 603
Qy 788 KCPSGTYGYGCRQICDCLNSTDHITG--TCYCSPGMRKANG 828
Db 604 SLDA-----CRSM-PCQNGATCVNNGADYVCECVGYAGQNC 639

RESULT 14
US-09-842-758-20
; Sequence 20, Application US/09842758
; Publication No. US20030083244A1
GENERAL INFORMATION:
APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R.
APPLICANT: Gerlach, Valerie
APPLICANT: Shinkets, Richard A
APPLICANT: Malysankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigaru, Muralidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esha A
APPLICANT: Smithson, Glenda
APPLICANT: Rastelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
FILE OF INVENTION: No. US20030083244A1
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758

```

```

; CURRENT FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/200,158
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: 60/200,613
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,780
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/201,006
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/201,007
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/201,236
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/201,238
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/201,186
; PRIOR FILING DATE: 2000-05-02
; PRIOR APPLICATION NUMBER: 60/201,474
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/201,508
; PRIOR FILING DATE: 2000-05-03
; PRIOR APPLICATION NUMBER: 60/220,591
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: 60/232,678
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 60/263,217
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: 60/265,160
; PRIOR FILING DATE: 2001-01-30
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 865
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-842-758-20

Query Match 12.0%; Score 808; DB 9; Length 865;
Best Local Similarity 33.1%; Pred. No. 2,2e-32;
Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

Qy 94 CCRPGYSESEMC-VHCA--DKCVHGR-CIANTQCCEFGMGKTSSACSDDHNGHCKT 149
Db 64 CCAGMRQGDDEGCIAGVEGNSVCSSENYCVRGECRGRGYGANCDDKCPROFGPPDK 123
Qy 150 SRCQCKNGALCPNITGACHCAAGFRGMRCEDEGQTYGNDCHORCCONGATCHVTGE 209
Db 124 ELCSCHPHQGCEDVYGGCTGHA--RRW-----GARCEHACQCOHG-TCHPRSGA 169
Qy 210 CCRPGYTGAFCEIDLCPPKHPQCEBQRCPCONGGVCHHVTGECSPSGMGTVCQGPCP 269
Db 170 CCEBGMWGA-----OCASACYCSATSRCDPOTGACLHAGW----- 206
Qy 270 EBRPFKNCSEQCGCHNGTCDAAATGCHCSPTGTGERODECPVGTGYLCAETGCVNG 329
Db 207 ---WGRSCNNQCAC--NSSPCEBQSGRCQCR---ER-----TFGARCDRTYQCBRG 249
Qy 330 GRCYHVSACLEAGFAGEARLCPEGLYIKDKRCPHLENTSHCPMSGECACKP 389
Db 250 -----RCHPYDGCACBP 262
Qy 390 GWSGLYCNETGSPFYGEACQOIC--SCONGADCDSVTGKC-TCADGFGIDCSTPCPLGT 447
Db 263 GYRGYKRCPCPAGFYGLCRRRCQKQOQCTVAEGRCILCEBGMWGTCKDPCANGF 322
Qy 448 VGINSSRC-GCKNAYVSPVDSCT-CKAGHGVDCSIRCSGTMGFGCNLTGCLNGS 505
Db 323 YEGGSHRCPPCRDHACHHVTGKTRCNAGNIGRCETKCSNGTYGEDCAFVCADCGSG 382
Qy 506 ACNTLDGTCTCAPGMRGEKCELPCQDGYGLNCAERCDSSHADGCHPTTGHC 557
Db 383 HDPFQSGKCLSPGYVHGPHCNVTCPGLHAGDACACSC-HDDTDPTYGAC 433

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:53:35 ; Search time 22.4565 Seconds

(without alignments)
1493.646 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVLINSLCFLICLLCHWI.....SSPKQEDSGSSSSSSSSSE 1140

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/PCFUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfill1a1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1037	15.4	2523	1 US-08-185-432-18	Sequence 18, Appl
2	1037	15.4	2523	4 US-08-899-232-3	Sequence 3, Appl
3	1035.5	15.4	2556	1 US-08-185-432-17	Sequence 17, Appl
4	1035.5	15.4	2556	4 US-08-899-232-2	Sequence 2, Appl
5	1034.5	15.3	2556	1 US-08-083-590A-20	Sequence 20, Appl
6	1034.5	15.3	2556	3 US-08-532-384-20	Sequence 20, Appl
7	1014.5	15.0	2471	1 US-08-185-432-16	Sequence 16, Appl
8	1014.5	15.0	2471	1 US-08-083-590A-19	Sequence 19, Appl
9	1014.5	15.0	2471	3 US-08-532-384-19	Sequence 19, Appl
10	1014.5	15.0	2471	4 US-08-899-232-1	Sequence 19, Appl
11	978.5	14.5	2703	1 US-08-185-432-19	Sequence 19, Appl
12	978.5	14.5	2703	4 US-08-899-232-4	Sequence 4, Appl
13	953.5	14.1	1964	4 US-09-467-997-1	Sequence 1, Appl
14	790	11.7	1053	4 US-09-214-278-2	Sequence 2, Appl
15	789	11.7	1212	4 US-09-214-278-3	Sequence 3, Appl
16	781	11.6	1257	3 US-08-611-729A-8	Sequence 8, Appl
17	779.5	11.6	1065	2 US-08-400-159-8	Sequence 8, Appl
18	779	11.6	1193	4 US-09-188-930-332	Sequence 332, App
19	779	11.6	1193	2 US-08-400-159-10	Sequence 10, Appl
20	779	11.6	1193	3 US-08-611-729A-10	Sequence 10, Appl
21	759	11.3	299	4 US-09-214-278-5	Sequence 5, Appl
22	757	11.2	1036	4 US-09-188-930-192	Sequence 192, App
23	757	11.2	1187	4 US-09-068-740A-6	Sequence 6, Appl
24	757	11.2	1187	4 US-09-068-740A-7	Sequence 7, Appl
25	757	11.2	1208	4 US-09-189-865-1	Sequence 1, Appl
26	757	11.2	1218	2 US-08-400-159-6	Sequence 6, Appl
27	757	11.2	1218	3 US-08-611-729A-6	Sequence 6, Appl

28	757	11.2	1218	4 US-08-882-046-2	Sequence 2, Appl
29	757	11.2	1218	4 US-09-068-740A-11	Sequence 11, Appl
30	755.5	11.2	2199	5 PCT-US95-11684-2	Sequence 2, Appl
31	755	11.2	1010	4 US-08-882-046-7	Sequence 7, Appl
32	755	11.2	1148	4 US-08-882-046-4	Sequence 4, Appl
33	747	11.1	1218	4 US-09-214-278-7	Sequence 7, Appl
34	745	11.0	1248	4 US-08-882-046-6	Sequence 6, Appl
35	744	11.0	1219	4 US-08-882-046-5	Sequence 5, Appl
36	709.5	10.5	1810	5 PCT-US95-11684-4	Sequence 4, Appl
37	698	10.3	3075	2 US-08-460-309-5	Sequence 5, Appl
38	698	10.3	3075	2 US-08-125-077-5	Sequence 5, Appl
39	690	10.2	1404	2 US-08-400-159-2	Sequence 2, Appl
40	690	10.2	1404	3 US-08-611-729A-2	Sequence 2, Appl
41	683	10.1	3111	2 US-08-460-309-4	Sequence 4, Appl
42	683	10.1	3111	2 US-08-125-077-4	Sequence 4, Appl
43	623	9.2	1417	4 US-08-900-230-3	Sequence 3, Appl
44	595.5	8.8	2211	4 US-09-738-884-1	Sequence 1, Appl
45	585	8.7	1345	2 US-08-977-767-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-185-432-18
: Sequence 18, Application US/08185432
: Patent No. 5750652
: GENERAL INFORMATION:
: APPLICANT: Artavanis-Tsakonas, Spyridon
: APPLICANT: Bussseau, Isabelle
: APPLICANT: Diederich, Robert J.
: APPLICANT: Xu, Tian
: APPLICANT: Matsuno, Kenji
: TITLE OF INVENTION: DETEIX PROTEINS, NUCLEIC ACIDS, AND
: TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PENNIE & EDMONDS
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/185,432
: FILING DATE: 21-JAN-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7326-006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2523 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-185-432-18
Query Match 15.4%; Score 1037; DB 1; Length 2523;
Best Local Similarity 25.4%; Pred. No. 1e-60;
Matches 326; Conservative 84; Mismatches 304; Indels 568; Gaps 78;
QY 83 GERTMYRR---KSCQ-----CP-GFYEGSCGVPHCAKCVHGR 117


```

Db 53 GERCOPNPCTIKNOCMNFETCEPVLQNAIDFICHCPVGF--TDKVCILPVDNACVNNP 110
  118 C-----IAPNTCOCEPEMGWGTNCSACDGDHMHCHTSRCKKNGALCNP--IT 164
  111 CRNGTCELLNSTYEXKRCRPPGWTGDSQQA-----DPCASN--PCANGKCLPFEIO 162
  165 GACHCAFRGRNCE--DRCEQ-----GTGNDCHOR-----CQ 196
  163 YICKPGRFHGATCKODINECSNPCKNGGQCINIEFGSTRCTQNFTGNDEPYPCN 222
  197 ---CONGATC--DHVTGRCRPPGYTGAFCED-----LC 225
  223 PSPCLNGGTCROTDDTSYCTCLPFGSGONCEENIDCPSNNCRNGTCVDGVTYNNCQ 282
  226 PRKHGPOCEQ---RC-----PCQNGVCHHTG--ECSGPGSMGTVCQ----- 266
  283 PDMWTQYCTEDVDECOLMPNACQNGGTCHNTYGYNCVNCWMTGEDSCSENIDCANNA 342
  267 -----PCPEGFRGKNC--SOEC---QCHNGTCD--ATGQ--CHCSPG 301
  343 CHSGATCHDRVASFYECCHGRGTLGLCHIDMNCISNPNCSNCDINPNVKAICTCPG 402
  302 YTGECQ---DECPVGTGYLCAETCQVNGGKCYHVSQA--CLCEAGAGERCEARLCP 356
  403 YTGPAKNNDVDECSIGAN-----PCRHGRCTYTLGSPQCNCPQYAGPRCEIDV-- 452
  357 EGLYIKCDKRC--PCHLHNTSHCHPMSE--CAKCPGMSGLYC-----396
  453 -----NECLSNPC--QNDSTCIDQIGEFQICMPYEGLYCETINIDECASNPCLN 501
  397 -----NE--TCSPGFYGEACQ---QICS---CONGADC-----DSYTK- 427
  502 GKCIDKINEFRCDCTPGFSGNLQOHDFDECTSTPCKNKAACLDGPNSTYCCQCTEGFTGRH 561
  428 -----CTCAPGPKGIDC-----STP-----442
  562 CEODINECTIPDCHYCTCKDGLATFTCLCRPGYTGRLCONDINECLSKPLNGGQCTDRE 621
  443 -----CPGLTYGINSR--CG---CKNDVACSPPVD--SCTKAGMHGVDCSIR---- 485
  622 NGYICCPRGTTGVNCEETIKIDCASNLCDNGKIDKIDGECTCEPGYTGKLCININIEC 661
  486 -----CBSGTWFGEC--495
  682 DSNPCNRNGTCKDOINGFTCVCPDGYDHMCLESEVNECSNPICIHACHDGVGNYKDCDE 741
  496 -----NLTCQ---CLNGACNTIDGT--CTCAPGMREKCEL-----PC-Q 530
  742 AGHSGSNCIDINNECESNFCOMNGTCKDMTGAYICTCKAGFSGPNCOJINIECSSNPCLN 801
  531 DGT-----YGLNC-----AERCD-----CSHADGCHPT--TGHCRLPGNS 564
  802 HGTCIDVAGYKNCMLPTGATCAEVLAPCAAGSPCKNGRCKESDPEFSECECPGNG 861
  565 GVHCD-----SYCABGRMBPNSL-----PCYCKNGASC 593
  862 GGTCEIDNECNVPRCNRNGATCONTNGSYKCNCKPGYTGRCNCEMDIDDCQPNCHNGSGS 921
  594 SPDDGI-----CECAPGFRGTGCR-----ICSPGFYGHNC 624
  922 S--DGINMFPCNCFRGPCKCEEDINECASNPCKNGANCTDCVNSTYCTCQDPGFSGIMC 979
  625 SOTCPQCVHSS---GPRCHITGL---CDCLPGFTGALC---NE-----658
  980 ESNTPCTESSCFNGTGC--IDGINTFQCCPFGTSGVCOQHDINCDKSPCLNGTQCD 1037
  659 -----VPSGRFRKNCAGI-----CTCTNNGTC--NPIDRSQCCTPGATGSSCSDP 702
  1038 SYGTGYKCTPOGTYGLNCONLYRMCDSPCKNGKCMQTNMFYR--CECKSGMTGYVCVDP 1096
  703 ---CPRA--HMGPNCHTCHNCHGARC--SAVDGCKCTPGMTGLCTCRCLPGLFGKNC 755

```

```

Db 1097 SVSCVAAKQGVDIYHL--CRNSGMCVDJGNTHFRCQAGTYGSYCEQV-----DEC 1148
  756 ALICOCONGADC-DHISG--QCTCRTFMGRHCBOK-----788
  1149 S-PNPOCNGATCTDYLGISCECVAGTGVNCSSEETNECLSHPCQNGGTCIDILINTYKCS 1207
  789 CPSTGYGCRQICD-----CLNNSTC-DHITG--TCYCSGPMKARGCDQAG 832
  1208 CPGRTGVHCEINVDCTPFYDSFTLEPKCFNNKIKIDYAGVGNJCICPPFVGERCE--- 1264
  833 VIIVGNLNS-LSRTSTALPADS 853
  1265 ---GDVNECLSN-----PCDS 1277

RESULT 2
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: OI, Huilin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

Query Match      15.4%; Score 1037; DB 4; Length 2523;
Best Local Similarity 25.4%; Pred. No. 1e-60; Indels 566; Gaps 78;
Matches 326; Conservative 84; Mismatches 304;

  83 GERTMYRR--KSQC-----CP-GYESGEMCVPHCADKCVHGR 117
  53 GERCOPNPCTIKNOCMNFETCEPVLQNAIDFICHCPVGF--TDKVCILPVDNACVNNP 110
  118 C-----IAPNTCOCEPEMGWGTNCSACDGDHMHCHTSRCKKNGALCNP--IT 164
  111 CRNGTCELLNSTYEXKRCRPPGWTGDSQQA-----DPCASN--PCANGKCLPFEIO 162
  165 GACHCAFRGRNCE--DRCEQ-----GTGNDCHOR-----CQ 196
  163 YICKPGRFHGATCKODINECSNPCKNGGQCINIEFGSTRCTQNFTGNDEPYPCN 222
  197 ---CONGATC--DHVTGRCRPPGYTGAFCED-----LC 225
  223 PSPCLNGGTCROTDDTSYCTCLPFGSGONCEENIDCPSNNCRNGTCVDGVTYNNCQ 282
  226 PRKHGPOCEQ---RC-----PCQNGVCHHTG--ECSGPGSMGTVCQ----- 266
  283 PDMWTQYCTEDVDECOLMPNACQNGGTCHNTYGYNCVNCWMTGEDSCSENIDCANNA 342
  267 -----PCPEGFRGKNC--SOEC---QCHNGTCD--ATGQ--CHCSPG 301
  343 CHSGATCHDRVASFYECCHGRGTLGLCHIDMNCISNPNCSNCDINPNVKAICTCPG 402
  302 YTGECQ---DECPVGTGYLCAETCQVNGGKCYHVSQA--CLCEAGAGERCEARLCP 356
  403 YTGPAKNNDVDECSIGAN-----PCRHGRCTYTLGSPQCNCPQYAGPRCEIDV-- 452
  357 EGLYIKCDKRC--PCHLHNTSHCHPMSE--CAKCPGMSGLYC-----396
  453 -----NECLSNPC--QNDSTCIDQIGEFQICMPYEGLYCETINIDECASNPCLN 501
  397 -----NE--TCSPGFYGEACQ---QICS---CONGADC-----DSYTK- 427
  502 GKCIDKINEFRCDCTPGFSGNLQOHDFDECTSTPCKNKAACLDGPNSTYCCQCTEGFTGRH 561

```

```

QY 428 -----CTCAGPKGIDC-----STP-----442
Db 562 CEEODINECTIPDPCHYCTCKDGIATFTCTLCRPGTGLCNINDINECTSKLNGOCTDRE 621
QY 443 -----CPLTGYVINCSSR---CG---CKNDWCVSPDC-SCYCKAGMHGVDOSIR-----485
Db 622 NGYICCPRGTTGVNCFETIDIDCASMLCNGKCIDKIDIECTCEGEGYTKLCNININC 681
QY 486 -----CPSGTWGFGC-----495
Db 682 DSNPCNNGSTCKDQINGFTCVCPDGYHDMCLSEVNECNPNICHGACHDGVGYKCDCE 741
QY 496 -----NLTCQ---CLNGACNTLIDGT---CTCARGMNGEREL-----PC-Q 530
Db 742 AGHSGSNCDIINNNECESNPNOMNGTCKDMTGATICTCKAGFSGPNCQTNINECSSNPNCLN 801
QY 531 DGT-----YGLNC-----AERCD-----CSHADGCHPT---TGHCRCLPGMS 564
Db 802 HGTICIDVAGYKCNMLPYTGAICCAVLAPCAGSPCKNGRCKESEDFTFSECECPGM 861
QY 565 GHYCD-----SYCABGRMGPNCSL-----PCYCKNGASC 593
Db 862 GGTCEIDMNECVNRCRNATCQNTNGSYKCNCKPGYGRNCEMDIDCQPNPCNHNGSGC 921
QY 594 SPDDGI-----CECAPGFRGTTCOR-----ICSPGFYGHRC 624
Db 922 S--DGIMNFCNCPAGFRGPKCEEDINECASNPCKNGANCTDVCNSTYTCTCOPGFSGIMC 979
QY 625 SOTCPGCVHSS---GPRCHITGL---CDCLPFGTALC---NE-----658
Db 980 ESNTPCTESSCENGGTC---IDGINTFCQCPGFGTGYCQNDINCDKSKPCLNGTQCD 1037
QY 659 -----VPSGFRGNKAGI---CTCTNNGTC---NPIDRSQCTPWTGLCTORCPLGFGSCQP 702
Db 1038 STGYTCTCPGQGTGLCNLVRMCDSSPCKNGKCKMOTNNEYR-CECKSGMTGYCQDVP 1096
QY 703 ---CPRA---HMGPNCIHTNCNHNAGFC---SAYDCECKCTPWTGLCTORCPLGFGSCQP 755
Db 1097 SVSCEFAAQQGVDIYHL---CRNSGMCVDTGNTHFCRCAGTIGSTICEQV-----DEC 1148
QY 756 ALICQCONGADC-DHISG-OCTCCTRTGFMGRHCEOK-----788
Db 1149 S--PNPQONATCTDYLGSGCEBAGYHGVNCESEINECLSHPCQNGTCTCIDLINTYKCS 1207
QY 789 CSBGTGYGCRQICD-----CLNNTC-DHITG-TYCYSGMNGARCDQXG 832
Db 1208 CPBGTGVCHEINVDCTPTFYDSFTLEPKCFNNGKCIDRVGGYNCICPPGFVEGRCE--- 1264
QY 833 VIIVGNLNS-LSRTSTALPADS 853
Db 1265 ----GDVNECLSN-----PCDS 1277

```

RESULT 3
 US-08-185-432-17
 Sequence 17, Application US/08185432
 Patent No. 5750652

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, Spyridon
 APPLICANT: Bussseau, Isabelle
 APPLICANT: Diederich, Robert J.

APPLICANT: Xu, Tian

APPLICANT: Matsuno, Kenji
 TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
 TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

```

; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-185-432-17

Query Match 15.4%; Score 1035.5; DB 1; Length 2556;
Best Local Similarity 25.9%; Pred. No. 1.3e-60;
Matches 317; Conservative 84; Mismatches 304; Indels 519; Gaps 74;

QY 94 CCGFYSEGMCVPHCADKRVHGR-----IAPNTCCQEPMGSTNGSSACDGDH 143
Db 89 CALGF---SGPLDITPLDNACLTNPNCRNGTCDLITLTKKCRPMSKSKQDA-----141
QY 144 WBPCHTSRCQCKNGALCNPTIGA---CHCAGFRG---WRCEDRCGEG---TYGNDCHQ- 193
Db 142 ---DPCASN-PCANGGQCLPFEASYICHCPSEFGPTCMQDVNECGQKPLTCHRGTCNHE 198
QY 194 ----RC-----OCONATC---DHVTEGECRPGYTGARPE-----222
Db 199 VGSYRCVCRATHTGPNCEMPYVPCSPSCQNGTCTPTGDTVTHECACLPGFTGQNCENI 258
QY 223 DLCPRG---KHGQC-----EORCP-----CONGVCHHVTG- 251
Db 259 DDCPGNCKNGACVGVNTYNCPCRPBMTGYQCTEDVDEQCLMPACQNGGTCHHTHG 318
QY 252 -ECSCPSSGMGTVCQ-----PCPEGRFGKNC--SOEC-- 281
Db 319 YNCVGVNWTGDEDCSENIDDCASACFHGATGCHDRVASFYCCBPHGRTGLCHLNDACIS 378
QY 282 -QCHNGGTDA--ATGQ---CHCSPGTGERCQ---DECFVGYGYVLCATCQCVNGKCY 333
Db 379 NPCNEGNSCDTNPVNGKALCTCPSGYTGACQDQDVDECSLGN-----PCEHAGKCI 430
QY 334 HTSGA---CLCEAGFAGERCEARLCPGLGKICDKRC---PHLEHTSCHPMSE--CA 386
Db 431 NTLGSECCQCLDGYTGPKEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
QY 387 CKPGMSGLYC-----NE---TCSPEFYGAQC---OICS--C 415
Db 478 CMRGYEGVCEVNTDECASSPCLHNGRCLDKINEFCCECPFTGHLQCYDVEDCASFP 537
QY 416 QNGADC-----DSV-TGKTCAPRFG 436
Db 538 KNGACCLDNPNTYTCVTEGTYGTGCEVDIDECDDPCHYGSCCKGVATFTLCRFGYTG 597
QY 437 IDCST-----PCPL-----GYGINS-----SRCCKNDAYS 465
Db 598 HICETININCSSQPCRLMTGTCDDPDNAYLCPGLKGTGTGNCINDDDCASSPDSGTCLD 657
QY 466 PYDG--SCYCKAGMHGVDCSIR-----CPSGTWFGGCL-----TC--- 499
Db 658 KIDGYEACAPGTYTSMCNINDECAGNPCNNGTCEIDINGFTCRCPGTYHDPICLSY 717

```

```

OY 500 -----OCLNGACNTLDG-TCTCAPRGKCEKEL-----PCODGTGLNC 538
DB 718 NECSNPPCVHAGCWDLSNKGKDCDDPGWSGTNDINNNECESNPCVNGCTCKDMTSGIYC 777
OY 539 A-----ERCC---DCSHADG-HPTTGH-CRCLPGWSGVHDSV---CAEG- 575
DB 778 TCWEGFSGPNQNTNINECASNPCLNKGTCTIDVAGYKCNCLLPYTATCEVYLAPCAPSP 837
OY 576 -RMGPNC-----SLPVC---KNGASCSPDDGICECAPRGFTTCORT-----CS 616
DB 838 CRNGGCRROSEDEYESFSCVCPAGAKGOTCEVDINECVLSPCMHGAASCONTHGXRYCHQ 897
OY 617 PGFYGHRSQTCPOC-----VHSSGPRHN---ITGLCDLPFTGALCNE----- 658
DB 898 AGYSGRNCETDIDDCMPNCHNGSGCTDGINFAFCDLPGFTGCEEDINECASDPCRN 957
OY 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCOCYPGWI 695
DB 958 GANCTDCVDSYTCCTCPAGFSGICHENNTPDCTESSCFNGGTC---VDGINSFTCLCPRGFT 1015
OY 696 GSDC-----SOP-----CPRAHGPNC---IHTCN---CHNGAFC 724
DB 1016 GSYCYHVNNECDSPCLLGCTGQDGRGLHRTCTPGYTGPNQNLVHWCDSPPCKNGKRC 1075
OY 725 ----SAYDGECKCTPGMTGLYCTQ-----R 745
DB 1076 WQTHQY---RCBCPSGWTGLYCDVPVSCEVNAORQGVDAVRLQHGGLCVDAAGTTHCR 1133
OY 746 CPLEFYGKDCALI---CQ---CONGADC-DHISG-OCTCRFGMRHCEOK----- 788
DB 1134 COAGYTGSYCEDLVDECSPPSPQONGATCTDYLGYSCKVAGYHGVNSEIDECLSHPC 1193
OY 789 -----CPSGTYGCRQICD-----CLNNSYC-DHITG-T 816
DB 1194 ONGGTCLDLPNTYKSCSCWGTQGVHCEINVDNCPVPDVPVSWSPKCFENNGTCVDQVGGYS 1253
OY 817 CYSPGKMGKARCQDQAGVIIVGNLN 840
DB 1254 CTCPPGFEVERCE-----GDVN 1270

RESULT 4
US-08-899-232-2
: Sequence 2, Application US/08899232
: Patent No. 6436650
: GENERAL INFORMATION:
: APPLICANT: Artavians-Tsakonas, Spyridon
: TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
: FILE REFERENCE: 7326-046
: CURRENT APPLICATION NUMBER: US/08/899,232
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 2556
: TYPE: PRT
: ORGANISM: Homo sapiens
US-08-899-232-2

Query Match 15.48; Score 1035.5; DB 4; Length 2556;
Best Local Similarity 25.9%; Pred. No. 1.3e-60;
Matches 317; Conservative 84; Mismatches 304; Indels 519; Gaps 74;

OY 94 CCGEYSGEMCVPHCADKCVHRC-----IAPNTCCQEPGNGTNCSSACDGDH 143
DB 89 CALGEF--SGPLCTPLDNLNACLTNCRNGGTCDLTLTLEYKRCRCPGMSGRSCQDA----- 141
OY 144 WGPHTSRCCCKNGALCNPTTGA--CHCAAGFRG--WRCEDECEG-----TYGNDCHQ- 193
DB 142 --DPCASN-PCANGGGLCPFEASTYICHCPSTFHTGTCWQDVNECGQAPRLCRHGGTCHNE 198

OY 194 -----RC-----OCONATC---DHVTGECKRPGYTGAFC----- 222
DB 199 VASRYCVCRATHGTGPNCEMPYPSPSPCONGATCRPTGDTVHCEACLPFGTQNEENI 258
OY 223 DLCPRG--KHGPOC-----EQRCR-----CNGVCHHTV- 251
DB 259 DDCPGNNCKNGACVDGVNTYKPCRPPEVTGQCTEDVDECOLMPNACONGTCHHTHG 318
OY 252 -ECSPSGMGTGYCQ-----PCPGRFRKNC--SOEC-- 281
DB 319 YNCVYVNGWTGDEDCSEINIDCASAOFHATGCHDVASTYCCSPHGRGLCHLNDACIS 378
OY 282 -QCHNGGTDA--ANGQ--CHCSPGYTGERCO--DECPVTTYVLAETCCGVNGKCY 333
DB 379 NPCNNGSNCDTNPVNGKALCTCPGSGYTGPACSDVDCELSGAN-----PCEHAKCI 430
OY 334 HVSGA--CLCEAGFAGEGRCARLCEGLYGLICDKRC--PCHLENTSCHHMSSE--CA 386
DB 431 NTLGSFECQCLQGYTGPCEIDY-----NECVSNPC--ONDAATCLDQIGEFQCM 477
OY 387 CKPGWSGLYC-----NE---TCSPGFYGEACQ---QICS---C 415
DB 478 CMPEGEVHCEVNTDECASSPCLHNGRCLDKINEPQCCPFTGHLQYDVBECASITPC 537
OY 416 ONGADC-----DSV-TGKCTCAPRGK 436
DB 538 KNGAKCLDGPNTYTCVCTEGYTGHCVEVDIDECDDPCHYSGCKXGDVATFTCLCPRYTG 597
OY 437 IDCST-----PCPL-----GTGYGNS-----SRGCKNDVNC 465
DB 598 HHCEININESSQPCRLMGTQODPNAYLCLCLKTTTPKNEINLDDCASSPDSGTCLD 657
OY 466 PVDG-SCYCKAGMHGVDNISIR-----CPSGTGFGCNL-----TC----- 499
DB 658 KIDGECACEGEGYTGSMCNSINIDECAGNPNCHNGTCEDEGINGFRCRCEGYHDPCLSEV 717
OY 500 -----OCLNGACNTLDG-TCTCAPRGKCEKEL-----PCODGTGLNC 538
DB 718 NECSNPPCVHAGCWDLSNKGKDCDDPGWSGTNDINNNECESNPCVNGCTCKDMTSGIYC 777
OY 539 A-----ERCC---DCSHADG-HPTTGH-CRCLPGWSGVHDSV---CAEG- 575
DB 778 TCWEGFSGPNQNTNINECASNPCLNKGTCTIDVAGYKCNCLLPYTATCEVYLAPCAPSP 837
OY 576 -RMGPNC-----SLPVC---KNGASCSPDDGICECAPRGFTTCORT-----CS 616
DB 838 CRNGGCRROSEDEYESFSCVCPAGAKGOTCEVDINECVLSPCMHGAASCONTHGXRYCHQ 897
OY 617 PGFYGHRSQTCPOC-----VHSSGPRHN---ITGLCDLPFTGALCNE----- 658
DB 898 AGYSGRNCETDIDDCMPNCHNGSGCTDGINFAFCDLPGFTGCEEDINECASDPCRN 957
OY 659 -----VCPSGRFGKNCAG---ICT---CTNNGTCNPIDR---SCOCYPGWI 695
DB 958 GANCTDCVDSYTCCTCPAGFSGICHENNTPDCTESSCFNGGTC---VDGINSFTCLCPRGFT 1015
OY 696 GSDC-----SOP-----CPRAHGPNC---IHTCN---CHNGAFC 724
DB 1016 GSYCYHVNNECDSPCLLGCTGQDGRGLHRTCTPGYTGPNQNLVHWCDSPPCKNGKRC 1075
OY 725 ----SAYDGECKCTPGMTGLYCTQ-----R 745
DB 1076 WQTHQY---RCBCPSGWTGLYCDVPVSCEVNAORQGVDAVRLQHGGLCVDAAGTTHCR 1133
OY 746 CPLEFYGKDCALI---CQ---CONGADC-DHISG-OCTCRFGMRHCEOK----- 788
DB 1134 COAGYTGSYCEDLVDECSPPSPQONGATCTDYLGYSCKVAGYHGVNSEIDECLSHPC 1193
OY 789 -----CPSGTYGCRQICD-----CLNNSYC-DHITG-T 816
DB 1194 ONGGTCLDLPNTYKSCSCWGTQGVHCEINVDNCPVPDVPVSWSPKCFENNGTCVDQVGGYS 1253
OY 817 CYSPGKMGKARCQDQAGVIIVGNLN 840

```

Db 1254 CTCPPGFVGERCE-----GDVN 1270

RESULT 5

US-08-083-590A-20

Sequence 20, Application US/08083590A

Patent No. 5786158

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/083.590A

FILING DATE: 25-JUN-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18.872

TELEPHONE/DOCKET NUMBER: 7326-015

TELEPHONE: 212 790-9090

TELEFAX: 212 8698864/9741

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 256 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-083-590A-20

Query Match 15.38; Score 1034.5; DB 1; Length 2556;

Best Local Similarity 25.88; Pred. No. 1.5e-60;

Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

Db 94 CCGPGESEGCVPFHCACKCVHGR-----IAPNTCCCEGSGGKGTSCSACDDH 143

Db 89 CALGE--SGPLCLPLDLNACLPNCRNGTCDLTLTXYKCRCPGSGKSCQA----- 141

Db 144 WCPHTSRCCCKNGALCPITGA--CHCAAGFRGRCE---DRCEG---TYGNDCHQ- 193

Db 142 --DPCASN--PCANGGCLPFPASTYCHCPSPFHGPTCKQDVNCCGQKRLCRHGSTCHNE 198

Db 194 ----RC-----OCONGATC---DHYWGECRCPPGTGAFC--- 222

Db 199 VGSYRCVCRATHGTGNCERPRYPVPCSPCONGTGRTGDTVHEACALPFTGQCEINI 258

Db 223 DLCPRG--KHGRQC-----EGRCP-----CONGVCHHTVG- 251

Db 259 DDCPPNNCKNGACVGVNTYNCPCRPMTGQYCTEDVDECOLMNAQNGSTCHNTGG 318

Db 252 -ECSCPSGMMGTVCQ-----PCPEGRGKNC--SQEC- 281

Db 319 YNCVAVNWTGDECSENIDDCASACFGATCHDRVASFYCECPRGKGLCHLDACTIS 378

Db 282 -QCHNGGTGDA--ATGQ--CHCSPGYTGERCO---DECPVGTGVLCAETCCQVNGKCY 333

Db 379 NPCNCSNCDTNPVNGKAICTCPSGYTGFPACSDVDDECSLGAN-----PCEHAGKCI 430

Db 334 HVSGA--CLCAGFAGECEARLCPREGLYGKDKRC---PCHLEHTSCHPMSGE--CA 386

Db 431 NTLGSFECOCLOGYTPRCEIDV-----NECVSNPC--QONATCIDQIGEFQCM 477

Db 387 CKPGMSGLYC-----NE---TCSDFGEACQ---QICS---C 415

Db 478 CMPTGEVHCEVNTDECASSPCLNANGRCLDXINERQCECPGTGTHLCOYDVDECASTPC 537

Db 416 QNGADC-----DSV--TGKCTCAPGFKG 436

Db 538 KNGAKCLDGPNTYTCVETGTHCEYDIDECDDPCHYSGCKDGVATFCLCRPGYTG 597

Db 437 IDCST-----PCPL-----GTGINS-----SRGCKNDAYCS 465

Db 598 HHCEININESSOPCRLRGTCODPDNAYLCLAGTGPNCINMDDCASSPCSGTCLD 657

Db 466 PVDG--SCTCKAGWGVDCSIR-----CPSGTWGFGCNL-----TC----- 499

Db 658 KIDYTECACPEGYTGSNCSNINDECAGNPCNNGTCEGJNGFTCRCPGEYHDPCLSEV 717

Db 500 ----OCLNGACNTLDG--TCTCAPGWRGKCEL----- 527

Db 718 NECSNCPVHACRDLNGLYKCDPDMGSGTNCIDINNECSNPNVCNGTCKDMTSGLYC 777

Db 528 PCQDSTYGLNCAERCD-----CSHADGC--HPTTGH--CRCLPGMSGVCHDSV---CABG- 575

Db 778 TCRBFGSPNCOITNINECASNPCLNKGTGIDVAGYKCNCLPYTGACVEVYLAFCAPSP 837

Db 576 -RMGPNC-----SLPCYC---KNGASCSPDDGICECAPGFRGTGORT-----CS 616

Db 838 CRNGECQSEDYNSFCVCTAGAKGQICEVDINECVLSFCRHGASCQNHGRTGRCQ 897

Db 617 PGFYGHRSQYCPQCVHSSGPCNH-----ITGLCDLPGFTGALCNE----- 658

Db 898 AGYSGRNCTETDIDC--RPNPCNHNGSCTDGINTFACDCLPGFRGTFCEDDINECASDPC 955

Db 659 ----VCSGSRFGKNCAG--ICT---CTNNGTCTNPIDR---SCQCLPG 693

Db 956 RANGANTCDVDSYCTCTCPAGSGIHCENNTPDCTESSCFNGGTC--VDGINSFTCLCPG 1013

Db 694 WIGSDC-----SOP-----CPAHMGPNC---IHTCN---CHNGA 722

Db 1014 FTGSTCOHYVNECBSPRCLLGGTCDGRGLRCHTCTPOSTYTPNCONLVHMDSSPCKNKG 1073

Db 723 FC---SAYDECKCTPGMTGLYCTQ----- 744

Db 1074 KCMQTHHTGY--RCBPSCGWTGLYCDVPSEVCEVAARQGVAVARLQHGGLCYDAGNTHH 1131

Db 745 -RCPGLFGYKCALI---CQ---CONGADC--DHSIG--OCTRTGPMGNHCEQK----- 788

Db 1132 CRCQAGYTSYCELDLVDECSPPCONGATCTDYLAGYCKCVAGYHGVNCESEEDICLSH 1191

Db 789 -----CPSGTYGCGRQICD-----CLNNSCT--DHITG 815

Db 1192 PCQNGTCLDLPNTYKCCCPRGTOGVHCEINVDNCPYDVYSNPKCFNNGTVDQYGG 1251

Db 816 -TCYCSPGMGKARCDQAGVITVGLN 840

Db 1252 YSCTCPPEFVGERCE-----GDVN 1270

RESULT 6

US-08-532-384-20

Sequence 20, Application US/08532384

Patent No. 6083904

GENERAL INFORMATION:

APPLICANT: Artavanis-Tsakonas, S. et al.

TITLE OF INVENTION: Therapeutic And Diagnostic Methods

TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And

NUMBER OF SEQUENCES: 21

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-20

Query Match      15.3%; Score 1034.5; DB 3; Length 2556;
Best Local Similarity 25.8%; Pred. No.: 1.5e-60;
Matches 316; Conservative 83; Mismatches 304; Indels 523; Gaps 73;

QY 94 CCGFYESEMCVPHCADKCVHGR-----IAPTCOCPEMGNGTSSACDGH 143
DB 89 CALGF--SGPLCLTPDMNCLTPRCNNGCTDLTLTEYKCRPFMSKSCQA----- 141
QY 144 WEPHCTSRQCKNGALCNITGA--CHCAAGFERGRC---DRCEG---TYGNDCHO- 193
DB 142 --DPCASN--PCANGGQCLPFEASAYICHPPSEHGPTCRDYNCGQKPRLCRHGTCHE 198
QY 194 -----RC-----OCQNGATC---DHVTEGECRCPGTYGAFCE--- 222
DB 199 VGSYRCVCRATHTGPNCRPRYPVPCSSPCONGTCRPTGVDVHEACLPFGTQNCENI 258
QY 223 DLCPFG--KHGPC-----EQRCP-----CONGVCHHTG- 251
DB 259 DDCPGNCKANGACVNDGVVTYNCPCRPMTGYCTEDVDECOLMAPACONGTCHHTHG 318
QY 252 -ECSCPSGMGTVCQ-----PCEGRFGKNC--SQEC-- 281
DB 319 YNCVAVNGMTGDECSINIDCASACFHGATCHDRVASFYCECPHRTGILCHLNDACIS 378
QY 282 -QCHNGTDA--ATGQ--CHSPGTGERQ---DECYVGYGVCAETCCQVNGCKCY 333
DB 379 NPCNESNCDTNPVNAKATCTCPSGTTPACSDVDVDECSIGAN-----PCHAGKCI 430
QY 334 HVSGA--CLCEAFAGEPEREARLCPGLVGIKDKRC--PCHLENTSHCHPMSE--CA 386
DB 431 NTLGSECCCLDGGTGPREFIDV-----NECVNPPC--QNDATCLDIDGFEQK 477
QY 387 CRPGMSGLYC-----NE--TCSPGFYGRACQ--QICS--C 415
DB 478 CMRGYGVHCEVNTDECASSPCLHNGRCLDKINEFCCECPPTGTHLCLQYDVEGASTPC 537
QY 416 QNGADC-----DSV--TGKCTCAPGFKG 436

```

```

DB 538 KNGAKLDPNNTYTCVTEGYTGHCEVDIDECDDPPCHYGSKDVAFTCLCRPGYTG 597
QY 437 IDGST-----PCPL-----GYGINS-----SRGCKNDAVVS 465
DB 598 HHETINECSSOPCRLRTGCTDPPDNAYLCLFCLKGTGPNCEINLDDCASSPDGCTCID 657
QY 466 PYDG--SCTCKAGHWGVDCSIR-----CPSGTWFGCNL-----TC----- 499
DB 658 KIDGYCACPEPYTGSMKNSNIDECAGNPNCHNGCTEDDINFTCCPBGHDPYCLSEV 717
QY 500 -----QCLNGACNTLDG--TCTCAPGMRGKCEL----- 527
DB 718 NECNSNPVCHGACRDLNGYKCDPCDPMGSGTNDINNNECESNPCVNGCTCKDMTSGIYC 777
QY 528 PDDGTGYNCAEROD-----CSHADG--HPTG--CHCLPGMSVHDSV--CAEG- 575
DB 778 TCHGEFSGPNQOTININECASNPCLNKGCTCIDVAGYKCKCLPBYGATCEVVLAPCAPSP 837
QY 576 -RWGPN-----SLPCYC--KNGASCSPDDGICECAPFRGTYTCORI-----CS 616
DB 838 CRNGGECRGSDEYSEFSCVCPAGAKGTCCEVDINCEVLSPCRHGASCONTHGYRCHQ 897
QY 617 PGFYGHRCQOTPCQYHSSGPFCH-----ITGLCDLPGFTALCNE----- 658
DB 898 AGYSGRNCETDIDDC--RPNPCHNGSGCTDGINTAFCIDLPGFRGTFCBEDINECASDPC 955
QY 659 -----VCPGFRGKNCAG--ICT--CTNNGTCNFIIDR-----SCQCYPG 693
DB 956 RNCANCTDGVDTCTCPAGFSGHCENNTPTCTESCTCNGGTC--VDSINSFTCLCPRG 1013
QY 694 WIGSDC-----SQP-----CPRAHWGPN--IHTCN--CHNCA 722
DB 1014 FTGSGYOHVYNECDSPRLGCTGQDGRGLHRCCTCPGTYGPNQNLVHMCSSPCKNGG 1073
QY 723 FC-----SAYDGECKTPRGTGLYCTQ----- 744
DB 1074 KCMQHTQY--RCECPSGMTGLXCDVPSYSCVAAORGVDAVRLCOHGSLCVDAGNTH 1131
QY 745 -KCPLEFGKDCALI--CQ--CONGADC-DHISG--QCTCRTFMGRHCEOK----- 788
DB 1132 CRQAGYTSYCEDVADYDESPSCQNGATCTDYGISCKYAGVNGVACSEBIDCLSH 1191
QY 789 -----CPSGTYGYGCRQID-----CLNNTC-DHTG 815
DB 1192 PCONGSTCLDLNPTYKSCSPRGTOGVHCEINVDNCPVDPVRSRSPCFNNGTCVDQVG 1251
QY 816 -TCYCSPGWKGARCDQAGYIYGNLN 840
DB 1252 YSCTCPGFGVERCE-----GDVN 1270

RESULT 7
US-08-185-432-16
Sequence 16, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavans-Teakonas, Spyridon
APPLICANT: Bussseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DETEYX PROTEINS, NUCLEIC ACIDS, AND
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

Query Match          15.0%; Score 1014.5; DB 1; Length 2471;
Best Local Similarity 23.0%; Pred. No. 3,1e-59;
Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

93 QCCPGESEBMCVPHCADKCVHGR-CIAPNTCO-----CEPGMGTCSSACDG 141
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
91 RCASGF--TGEDCOYSTSHCFVSRPCLNGTCHMLSRDYEECTQVGFGRKEC----- 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
142 DHMGPHSTSCQCKNGMLCPRTG--ACHAGRGRGRCDRCRCEQGYGDCIHRCCOON 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
143 -QMTDACLH-PCANGSTCTTANQFSCKCLTGTGKCE-----TDVNECDIPGHQH 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
200 GATCDHYTG--ECRCRPGYTGAFCEDL-----CPRGRHG 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
195 GGCCLNLPGSYCCQCPGFGFYGYCDLSLYPCAPSPCVNGGTGRTGDTFTECNLPRFEG 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 PCEBQ-----RCP-----CONGV 245
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
255 STGERNIDCPNHRCONGVGVVDGVTNYNCRCPQMTGQFCTEDVDECLLPNACONGGT 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
246 CHHYTG--ECCSPSGMGVTCGO-----PCBGRFGKNC- 277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
315 CARNNGYGCYCVNMGSGDDCSENIDCAFACTPGSTCIDRYASFCMCPBGRAGLLCH 374
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
278 -SQEC--OCHNGGTCA--ATGO--CHCSPGYTGERCO--DECPVGYGVLCATETCQC 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 LDACISNPGCHKGALCDOTNPLNGYICTCPQGYKAGADCTEDVDCCAM-----ANSNPC 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 VNGKCYHVGGA--CLCEAFAGEGEC-----ARLCPEGLY 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 EHAQKCVNTDGAHFCECLKAGPRCEMDINECHSDPCQDANCLDKIGGFTCLCMGFK 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 GIKCDKR-----CP-----CHLE-----NTHSC 378
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
488 GVHCELEINECQSNPCYNNQCCYDKVNRFOCLCPRGFTGVCOQIDIDDCSSTPLCLNACK 547
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
379 --HPMSGECACKPMGSLYCNET-----CSPGFYEAC-QQ 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
548 IDHNGYECOCATGFTVLEENIDNCDPRCHHGOCQDDISDYTCICNGYMGALISDQ 607
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
412 I-----CSCONGA-----DCDS--VTG-----K 427
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
608 IDECYSSPLCNDRCIDLVNGYOCNCPGTSGVNCIEINFDCASNPCIHICMDGINRYS 667
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
428 CTGAPGKSIDC-----STPCLGTGYNCS--RCGCK-----NDAYCS 465
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
668 CVCSPGTGRCNIDIDECASNPCKKATGTCINGVNGRCICPBGPHPSCTSOYVNECLSN 727
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 P-VDGSCCT-----CKAGHGVDCSI-----RCPSTGT 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

728 PCHHGNTGSLSGKCLCDAGWGWINGEYVKNNECLNPNCONGTCDNLVNGYRCTCKKE 787
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
491 WGFQCNLTQ-----CLNGAC----- 507
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
788 KGYNCQVINIDECASNPNCLNNGTCTCFDDISGYTCHVLPYTGKNCQTVLAPSPNCEAAV 847
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
508 -----NTLDTGTCAPGRMBREKELPCQDDTYGLNCAERDCSHADCHFTTGH--CRCL 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
848 CKSPNESTYCLCAPGMOGORTIDIDE-----CLSK-PCMHNGCHHTQSSYMCCEP 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
561 PGMSGYHCDSDVCAEGRMGPMSLPCYCKNGASCSPDDGI-----CECAPRGRTTCOR--- 613
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
901 PGSGMDCEEDIDDCLANP-----CQNGSCM--DGVNTFSCCLCPGTGDKCQTDNM 951
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
614 -----ICSPGYHGRSCQPCQVHSS--GPCHHITGL--- 644
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
952 ECLSEPCKNGTGSDYVNSYTKCQAGFDGVHCENNINECTESSCFNGTGC--VDGINSF 1009
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
645 -CDCLPFTGALC-----NEV-----CPSGRGRKNC--AGICT- 674
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1010 SCLCPVGFETGSCFLHETNECSSHPCLNETGCVDLGTYRCSCPGLGTGKNCQTVLNLCSR 1069
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
675 --CTNNGTC--NPIDRSCQCYPMGIGSDCSQP----- 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1070 SPCKNKGTCYQKKAESQCLCPSGMAGAYCDVPNVSCDIASRRGVLEHLCHQSHGVICNA 1129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
703 -----CPRAHMGPNC--IHTC--NCHNAPCSATYDG--BCKTPTGTLGYTOR-- 745
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1130 GNTHYCCPLGYTGSYCEBOLDPCASNPCOHGATCSDFIGYRCECPYGOGVNCEYEVD 1189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
746 -----CPLGFG-----KDCALICQONGADG--DHISGO- 773
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1190 ECONOPCONGTCTIDLVNHFKCSCPETRGILLCEINIDCARBPCLNGSCQCHDRIGYS 1249
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
774 -----CTCRTPGMRHCEQKCPSTGT 793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1250 CRCLPFAGERCBGDIINECLSNPCSSSEGLDCLQTLFNDYLCVCRSAFTGHRCE----- 1303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
794 YGIGCRQICDCLNNSTCDHTTG-----TYCSPGMRKARCDQAGVILVNLNLSRSTRA 848
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1304 FVDVCPDM--PCLNGGTCVAVASNMPDGFICRCPGFSSGARS--CGGVKC----- 1351
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
849 LPADSVQIGAIAGIILVLFLLALFIYRHKQKKE-----SSMPAVYTTTPAMRYVNA 904
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1352 -----RKGEQCVHTRASGPR-CFCSPRDCES 1376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
905 DYTISGTLPHSNGNMANSHYFTNPSTHTLQCATSPHVNRRDRTYTKSRNNOLEFVNLKN 964
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1377 GCASS--PCQHGSGC--HPQRPPTYXS-CQCA-PPFSGR-----CELYT 1415
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
965 VNGKRGVGDCTGTLPADMKHGGYLNELGAPGLDSYMKSLDKLAKNS---EYSSNC 1021
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1416 APPST--PPATYCLSYQCADARQVDCDE-----ACNSHACQMDGDC 1455
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 SLSSENRYATIKDP 1036
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1456 SL-TMENPANCSSP 1469
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Atavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penile & Edmonds
; STREET: 1155 Avenue of the Americas

```

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/083,590A
FILING DATE: 25-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Miarock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-083-590A-19

Query Match	15.0%	Score 1014.5	DB 1	Length 2471
Best Local Similarity	23.0%	Pred. No. 3.1e-59		
Matches 348	Conservative 106	Mismatches 354	Indels 707	Gaps 78

```

0Y 93 QCCGFESGEMCVPHADKCVHR-CIAPTCO-----CEGMEGJTNSSACDG 14.1
Db 91 RCASGF--TEGDQYSTRSHCPVSRPLNGSTCHMSLRDTYECTQYGFYKKE-----14.2
QY 142 DHMGPHCTSRQCKNGALCNPTNG--ACHCAAGFRGNRCEDRCEQGTGYNDCHORCOQN 19.9
Db 143 -QMTDACLSH-PCANGSTCTTVANQFSCKLTGTGKCE-----TDVNECDIPHCQH 19.4
QY 200 GATCDHVTG--ECGCPRGYTGAFEBD-----CPGKHG 23.1
Db 195 GGTGLNLPGSYQCCPGFTGQYCDLSYVPCAPSPCVNGGTCRGTGDTFTECNCLPFGEG 25.4
QY 232 PQCEQ-----RCP-----CONGV 24.5
Db 255 STCBRNIDDCNHNRCQNGVCYDVSVNTYNCRCRCPQMTGQCTEBDYDECLLDQPNACONGGT 31.4
QY 246 CHAVTG--ECSPSGMNGTVCGO-----PCPEGRFGKNC- 27.7
Db 315 CANNGGVCYGVNGMSGDSCDENIDCAFASTPGSTCIDRVASFSCMPGEMAGLCH 37.4
QY 278 -SQG---QCHNGSTCA--ATGQ--CHCSBGTGECQ--DECPYGTGYVLCALFTQCQ 32.6
Db 375 LDDACISNPCKRGALCTNPPLNGRYICTCPQYKGADCTEDVDECAM-----ANSNPC 42.7
QY 327 VNGKCHYVSGA--CIEAGPAGRCE-----ARLCEPGLY 36.0
Db 428 EHAKCVNTDGAFFHCECLKYAGFRCEMDINECHSDPCQNDATCLKTIGFTCLCMRFGK 48.7
QY 361 GIKCDKR-----CP-----CHLE-----NTHSC 37.8
Db 488 GVHCLELEINECQSNPCVNVNGCVDKVNRFQCLCPRFTGPVCOIIDDSCSTPOLNAGKC 54.7
QY 379 --HFMSECKACKPWSGLYCHET-----CSFGTYGEMAC-QQ 41.1
Db 548 IDHPNGYECOCATFTGVLCEENIDNCDPPCHHGQCODEIDSYTCICNPGYMAICSDQ 60.7
QY 412 I-----CSCONGA-----DCDS--YMG-----K 42.7
Db 608 IDECYSPCLMDGACIDLWNGYQCCPGYSGVNCCEINFDCAASNPCITHICMDGINRYS 66.7

```

0Y	428	CTAPGFGIDC-----SFPCLGTGYNSS--RCGCK-----NDAYS	465
0Y	468	CVCSSGFTGQRNCIDNIDECASNPCRKRGATCINCYNFRCICEGPHHPSCTSOVNECLSN	727
0Y	466	P-VDGSCF-----CRAGMHGVDCSI-----RCPSGT	490
0Y	728	PCIHGNTGGISGYKCLCAGAWGINCVEYDNKNECLSNPCQNGTGODNLVNGRCCKKGF	787
0Y	491	WFGGCLNCTCQ-----CLNGAC-----	507
0Y	788	KGYNCQVNIDECASNPCCLNGTGFEDDISGYTCHCJLPTYGKNCQTVLAPCSPNCENAV	847
0Y	508	-----NLLDGTCTAPGMBREKELCQDQSTYLNCAEKGDCSHDGGHPTTGH--CRCL	560
0Y	848	CRKSPNEBESYTCCLAPBQMQRCTIDIDE-----CISF-PCMNGLCLNTGOSTYMCCEP	900
0Y	561	PWMSVHCDSDVCAEGRMGPNCSTLPCYCKNGASCSPDGI-----CECAPFRTGTCOR--	613
0Y	901	PGFSGMCEEDIDCLANP-----CQNGSCSM--DGWTFSSCLCLEPFTGDKQOTDMN	951
0Y	614	-----ICSPGFYGHRCSTQCPOCVHSS---GPCHHITGL---	644
0Y	952	ECLSEPKNGTCSDYVNSYTKCCKAGFEDVHCENNINECTESSCFNGTGC--VDGINSF	1009
0Y	645	-CDCLPFGTALC--NEW-----CPSRFGKNC--AGICT- 674	
0Y	1010	SCCLPVGSTGFSFCHETNECSSHPCCLNEGTGVDGLTGYRCSPCLGTGKNCQTLVNLCSR	10659
0Y	675	-CTNNGTGC--NPIDRSCQCYCPGMIWSDCSQP-----	702
0Y	1070	SPCKNGKGYCKAKAESQCLPSPGMAGACDVPYVSCDDIASNRGVLVHILCOHSGVCINA	11229
0Y	703	-----CPRAHMGPNC---IHTC--NCHNGAFCSAYDG--ECKCTPQWTLGYCTOR--	745
0Y	1130	GNTHYCCPLGTYGTYSCYCEQDLEDCASNPCOHGATCSDFIGRGYRCCEVGYOGNCEYEVD	1189
0Y	746	-----CPJGFFYG-----KCCALICQCGANGAO--DHISQ- 773	
0Y	1190	ECQNOPCONGSTCIDLVNHFKCSCPGRTGRLCEENIDICAGPHCLNGGQCMDTIGYS	1249
0Y	774	-----CTCPTGFMGRHCEQKCPSGT 793	
0Y	1250	CRCLPFGAGEBCEDNIDECASNPCSSBSGLDICTQLNDVLCYCRARAFTRGRHE-----T 1303	
0Y	794	YGYGCRQICDCLNSTDHITG-----TCYCSFGMKGARCDOAGYIYGNLNSLSRISTA	848
0Y	1304	FVDVCPQM--PCLNGTCAVANSNMPDGFICRCPGFSGARCOS--CGOYVC-----	1351
0Y	849	LPADSYGAIAGIILIVLVFLMLFIYHRKQKGE-----SSMPAVTYTPAMRVYNA	904
0Y	1352	-----RKGEQCVHTFASGPR-CFSPSPRDECS 1376	
0Y	905	DYTIISGTLPHSNGNANSHFYFNPYSYHTLLOCATSPHVNRRDMATVTKSKNNOLVFNLKN	964
0Y	1377	GCASS--PCOHGSSC--HQQRQPRYYS-CQCA-PPFSSR-----CELYT 1415	
0Y	965	VNPGKRGVGDCTGTLPADMKHGGYINELJAGFSLDRSYKGLKDLGKNS--EYNSNC 1021	
0Y	1416	APPST--PPATCLSQYCADKARQVDCDE-----ACNSHACQMDGDC 1455	
0Y	1022	SLSSSENVPATIKDP 1036	
0Y	1456	SL-TWENPMANCSP 1469	

RESULT 9
 US-08-532-384-19
 ; Sequence 19, Application US/08552384
 ; Patent No. 6083804
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavals-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods

RESULT 10
 US-08-899-232-1
 ; Sequence 1, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILE REFERENCE: 7326-046
 ; CURRENT APPLICATION NUMBER: US/08/899,232
 ; CURRENT FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2471
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-08-899-232-1

Query Match 15.0%; Score 1014.5; DB 4; Length 2471;
 Best Local Similarity 23.0%; Pred. No. 3,1e-59;
 Matches 348; Conservative 106; Mismatches 354; Indels 707; Gaps 78;

Oy 93 OCCPGTESEGCVPBHCADKCVHGR-CIAPNTCO-----CEPGMGTCNSACDG 141
 Db 91 RCASGF--TGEDCOYSTSHPCFVSRLCLNGTCHMLSRDTYECTGOVFTGKEC----- 142
 Oy 142 DHMGHCHTSRCKCKNGALCNPTIG--ACHCAGFRGMRREDCEGGTYGNDCHORCOCON 199
 Db 143 -OWTDACLSH-PCANGSTCTTVANQFSCCKLTGTGQKE-----TDVNECDIPRHCQH 194
 Oy 200 GATCDHVTG--ECRCPGTYGAFCEDL-----CPGKKG 231
 Db 195 GGTCLNLPESYGCQCPGFTGTGYSLSYPCASPVCNGTGRGTGDTFTFECNCLPFGG 254
 Oy 232 POCEDQ-----RCP-----CONGV 245
 Db 255 STCEBNIDDCPNHRCONGCVGVNTYNCRCRPMQTGCTEDVEDCLLPNACONGST 314
 Oy 246 CHHVTG--ECSPSGMNGTVCGQ-----PCPRGRGKNC- 277
 Db 315 CANRNGIGCVNCGNSGDDCSENIDDCAFASCTPGSTCTIDRVASFSCMCPGKAGLICH 374
 Oy 278 -SQEC--QCHNGTQDA--ATGO--CHCSPGTGERCO---DECPVGTGVLAETQC 326
 Db 375 LDDACISNCHKGAICDTPRLNGOYICTPQGYKGAADCTEDVEDCAM-----ANSNRC 427
 Oy 327 VNGGACVHVSQA--CLCEAGFAGERCE-----ARLCEGLY 360
 Db 428 EHAGCAVNTDGAHFHCCLKLGAYGRCEMDINECHSDPCONDATCLDKIGFTCLMPGRK 487
 Oy 361 GIKCKR-----CP-----CHLE-----NTHSC 378
 Db 488 GVHCELEINECOSNFCVNVNGOCVDKYNRFCLCPRGFTGPVCOIDIDDCSSPTCLNGAKC 547
 Oy 379 --HPMSGBCAKPGWSGLYCNET-----CSPGFYBAC--Q 411
 Db 548 IDHPNGYEGOCATGFTGVLCSEINIDCDRPRCHGOCQDIDSYNICIPGTMGALCSQ 607
 Oy 412 I-----GSCQNGA-----DCDS--VTG-----K 427
 Db 608 IDECYSSPLNDRCIDLVNGYCNOCOPGTSGVNCSEINEDDCASNFCIHGICMDINRYS 667
 Oy 428 CTCAGFGKIDC-----SPPCLGTGICSS--RCGCK-----NDAYS 465
 Db 668 CVCSSGFTGQRCHIDIDECASNPKRKGATCINGVNGFRICICEGPHHPSCYQVNECSLN 727
 Oy 466 P-VDSCCT-----CKAGHGVDCST-----RCPSGT 490
 Db 728 PCHHNCNCGLSGKCLDADAGVNGINCVDKNECLSNPCQNGCTDNLVNGIRCTCKKGF 787

Oy 491 WFGCNLTQC-----CLNGAC----- 507
 Db 788 KGYNCQVNIIDECASNPCLNMGTCFDDISGYTCHCVLPYTKNCQTVLAPCSPNCEANAV 847
 Oy 508 -----NTLDGTTCAPRGKRGKCELPCCODGTGGLNCAERCCSHADGCHPTTGH--CICL 560
 Db 848 CKESPFEYSYTLCAFGMOGQCTIDIDE-----CISK-PCMNIGLCHNTOGSYMCCEP 900
 Oy 561 PGMSGVHCDSYVCAEERMGNCSLPCYCKNGASCSPDDGI-----CECARGFRGTTCOR-- 613
 Db 901 PGFSMGDEBEDIDDLANP-----CONGSCM--DGVNITSCICLPGFTGDKCOTDMN 951
 Oy 614 -----ICSPGFYGHRCQSOTCPQVHSS--GPCHHITGL-- 644
 Db 952 ECLSEPKKNGTCSPTYVNSYTKCOAGFGVYCNENINECTSSCFNGGTG--VDGINSF 1009
 Oy 645 -CDCLPGFTGALC-----NEV-----CPSRFKNC--AGICT- 674
 Db 1010 SCLCPVGFGTSPCLHEINECSSHPCLNBTGTCVDTGTYRCSCPLGTYRKNCTLVNLCR 1069
 Oy 675 --CTNNGTC--NPIDRSCOCYPGWIGSDCSP----- 702
 Db 1070 SPCKKKGTCVOKKASQCLCPSGMAGATCDVNVNSCDIAASRGVLEHLCOHSGVCINA 1129
 Oy 703 -----CPRAHMGPNC--IHTC--NCHNCAFCSAYDG--ECKCTPMTGLYCTOR-- 745
 Db 1130 GNTHYCQCPPLGYTGSYCEBQLEDCASNPQOHGATCSDFIGYRCBCVPGYQVNEVEVD 1189
 Oy 746 -----CPLGFTG-----KDCALICQCONCADC-DHISQ- 773
 Db 1190 ECOMQPCONGTCLDLVNHFKCSPPTGTLICEENIDDCACAGPRLNGQCMIDIGYS 1249
 Oy 774 -----CTCRPGFMGRHCEQRCPSGT 793
 Db 1290 CMCLEGFAGERCEGIDINECLSNPCSESSLDICQLTNDYICVCRAFTGRHE-----T 1303
 Oy 794 YGYGCRQICDCINSTCDHTG--TCYSGPKNKARQDAGVYIVGNLSLSRTSTA 848
 Db 1304 FVDVCPOM-PCJNGGTCAVANSNMPDGFICRCPPGFSGARCOSS--CGQVYC----- 1351
 Oy 849 LPADSYQIGALIGITIVLVFLALFTIYHNKQGE--SSMPAVTYPRAMRYNA 904
 Db 1352 -----RKGEQCHTASGPR-CFCPSRDCES 1376
 Oy 905 DYTISGTLPHSNGNANSHYFTNPSTHTLOCATSPHYVNNRDMRTVTSKNNQLFVNKN 964
 Db 1377 GCASS--PCQHGSC--HPQKQPPYYS-CQA-PPFSGR-----CELYT 1415
 Oy 965 VNPGRKPGVGDCTGTLPADMKHGGYLNELGAFGLDRSTWGSLSKDLGKNS--EYNSNC 1021
 Db 1416 APPST--PPATCLSQYCADKARDGVCD-----ACNSHACQMDGDC 1455
 Oy 1022 SLSSSENPTYATIKDP 1036
 Db 1456 SL-TWENPMANCCSP 1469

RESULT 11
 US-08-185-432-19
 ; Sequence 19, Application US/08185432
 ; Patent No. 5750652
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Busseau, Isabelle
 ; APPLICANT: Diederich, Robert J.
 ; APPLICANT: Xu, Tian
 ; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
 ; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: PENNIE & EDMONDS
 ; STREET: 1155 Avenue of the Americas

```

: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2711
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/185,432
: FILING DATE: 21-JAN-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7326-006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
:
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 19:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2703 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-185-432-19

```

Query Match 14.5%; Score 978.5; DB 1; Length 2703;
 Best Local Similarity 26.6%; Pred. No. 8,6e-57;
 Matches 309; Conservative 105; Mismatches 331; Indels 417; Gaps 72;

```

QY 3 ISINSLSTICLLCHWIG-----TASPLNLEDPNVC-SHWESYVTVOESYHPDP 53
DB 231 ICVNHGSIQCMCPGTYTKDCDITKPKSPSCQNAIGICRSNGLSYECKCPGEGKNC 290
QY 54 QIYYSCTDIL--NMFKCTRHRVSYRTAYRHEKTYRRKSOCPPGFYSEGMCC--VPH 108
DB 291 EGNVDDCLHLIDONGSTCIDGISDY-----TCKRPPNF--TGRRCODDVE 334
QY 109 CADK---CVHGRCIAPNT-----CQCEPQNGVYCH-HVYG-ECSCPSGMWGTVC---T 149
DB 335 CAQRHPVQNG-ATCTNTHGSIYSCICVNGMAGLDCSNNTDCKQNAEYATCIDGVGS 393
QY 150 SPCOC---TNGALC-----NPTG--ACHCAGFGMGCE---DGC 182
DB 394 FYCQCKGKGTGLCHLDACISNPHADAICTISPLNGSYACSCATGYRGVDCSEDDIC 453
QY 183 EGGTGNDCRQCCONGATCDHYTG--ECRCPPGYTGAFCED----- 223
DB 454 DQGS-----PCEHNICVNTPSYRNCNSQSGFTGRCETNINECSHPCQNGSCL 504
QY 224 -----LCPPKHPGCE-----QRCPCQNGVYCH-HVYG-ECSCPSGMWGTVC---T 264
DB 505 DDPGRFCVCMGFTGTQCEIDIDECQSNPCLINDTCHDKINGKSCALGFTGARCQIN 564
QY 265 -----GQF-----CPBGRFGKNC--QECQ--CHNGTCDATG-Q 295
DB 565 IDDCOSQPCRNNGICHDSIAGYSCBPPGYTGTSCBININDSDNCHGKCIDVNSFK 624
QY 296 CHCSPGYTERCO--DECP-----VGTGYVIC-----AETCO----- 325
DB 625 CLCDPRTYITICOKINDESNPCORPDGHCODRVGSYTCQCCAGTSGKNCVNVNCHSN 684
QY 326 -CVNGSKCYHVSQ---ACLCEGAFGERCEAR-----LCBEGLYGKCDKR 367
DB 685 PNNNGATC--IDGINSYKQCVGFTGHCCKVDECISSPCANNVCIIDVNGYKCE-- 740
QY 368 CQCHLENTH-----SCHPMSEGCACKRPMGSLCNETCSGFGFGEACQ---QICS-- 414
DB 741 CPGRYDAHCLSDVDECAASNPCVNEGRCEBDGINEFTCH--CPGYYGKRCCELDIDECSSN 798

```

```

QY 415 -CONGADC-DSVYG--KCTCAPGFKGIDCST-----PCPIG----- 446
DB 799 PCOHGCTCYDKLNAPSCQCMPTGTGCKCETINDDCTNCGNGNCTIDKYNKCYKVP 858
QY 447 TYGNCSSR--CG---CKNDVCSF---VDGCTCKAGHWGDCSIRCPSGTWGFGCN 496
DB 859 FGRDCESHMDPCARRNCKNEAKCTPSSNPLFSCYCKLGTYGRYDEDEDE-----CS 912
QY 497 LRCOCINGACMTLDT--CTCAPGMRGKCEL-----PQDDTYGLNCAERDCSH 546
DB 913 LSSPCRNASCLNVPSTYRCLCTKGEGRDCAINTDDCASFPQONR-----TCLD 963
QY 547 ADGCHPTTGCHRCCLPMSVGHSDVCAEGRMGPN-C-SLPYCKNAGASCSP--DDGICEA 603
DB 964 GIGDY-----SCLCYDFDCKHCETDNE-----CLSQP--CONGATCQYVNSTYCTP 1011
QY 604 PGRFTTCOR-----ICSPGFYGRCSQTPQCV--HSSGPCHHITGLCDC 647
DB 1012 LQFSGINCQTNDEDTRESSCLNGSGCIDGINGNCS-----CLAGSGANCOYKLNKCD 1066
QY 648 LQFGLCALCNEY-----CPSGRGKNCA-----GICTCNNGCNFIDR--SOOCYP 692
DB 1067 NPCLNATCHEQNNNEYTCHCPSPGFTGKCSEYVDMCGOSPCEGATCSQMKHOFCKCSA 1126
QY 693 GWIGSDCSQPCPBAHMGPCIHCHNCH-----NGAFCSAYDGE--CKCTPG 736
DB 1127 GWTGKICD-----VQISQDADARKGLSLRLCLNNGTCKDKDGNHVCYCSOG 1174
QY 737 WTLGYTCORCLPGLFYKDCALICQ---CONGADCHISG--OCTCTGFMGRHCEO---- 787
DB 1175 VAGSYCKEID-----EQSQPCONGGTCRDLIGAYECQCRQGOGNCLEINID 1224
QY 788 -----KQPSGTGYGC--RQID--CLUNSTC-DHITG--TCY 818
DB 1225 CAPNCPQNGCTCHDRYNAFSCSPETMGIICEINIKDKCPGACHNNGSCIDRVGSECV 1284
QY 819 CSPGMKARCDAQVLIIGNLN 840
DB 1285 CQPGFVARGC-----GDIN 1299

```

RESULT 12
 US-08-899-232-4
 ; Sequence 4, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Aftaknas, Spyridon
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILE REFERENCE: 7326-046
 ; CURRENT APPLICATION NUMBER: US/08/899,232
 ; CURRENT FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 4
 ; LENGTH: 2703
 ; TYPE: prt
 ; ORGANISM: Drosophila sp.
 ; US-08-899-232-4

Query Match 14.5%; Score 978.5; DB 4; Length 2703;
 Best Local Similarity 26.6%; Pred. No. 8,6e-57;
 Matches 309; Conservative 105; Mismatches 331; Indels 417; Gaps 72;

```

QY 3 ISINSLSTICLLCHWIG-----TASPLNLEDPNVC-SHWESYVTVOESYHPDP 53
DB 231 ICVNHGSIQCMCPGTYTKDCDITKPKSPSCQNAIGICRSNGLSYECKCPGEGKNC 290
QY 54 QIYYSCTDIL--NMFKCTRHRVSYRTAYRHEKTYRRKSOCPPGFYSEGMCC--VPH 108
DB 291 EGNVDDCLHLIDONGSTCIDGISDY-----TCKRPPNF--TGRRCODDVE 334
QY 109 CADK---CVHGRCIAPNT-----CQCEPQNGVYCH-HVYG-ECSCPSGMWGTVC---T 149

```

```

Db 335 CAORHPVQNG-ATCTNTHGSYSYCIVNGMAGLDCSNNTDOCKOACACGYATNCIDVGS 393
Oy 150 SNGQC---KNGALC-----NPTTG---ACHCAAGFGMWCE---DRC 182
Db 394 FTCCQCKGTGLCHLIDACTSNPCHADAICDTPSINGSYACSCAAYGKGVCSSEIDEC 453
Oy 183 EGGTGNDCHQKCCONGATCDHVTG--ECRCPPGYTGAFCED----- 223
Db 454 DQGS-----PCENHCICVNTPSYSTNCNSQGTGRCETNINECSHPCQNEGSCL 504
Oy 224 -----LCPPKKGPOCE-----QRCPCONGVCH-HVTG-ECSCPSGMAGTVC--- 264
Db 505 DDPTGFRVCYMPGFTGTCEIDIDECQSNPCINDGTCHDKINGFKSCALGFTGANCOIN 564
Oy 265 -----GDP-----CPBGRFKNCS---QEQO---CHNGTCDATG-Q 295
Db 565 IDDCSQSPRNRNGICHDSTAGYSCBPCPYGTGTSCENINDCSNPNCHRGKCIDVNSPK 624
Oy 236 CHCSPGYTGERCO---DECP-----VGTGVLC-----AETCO----- 325
Db 625 CLCDDGYTYICQKQINECSNPOCFDHCQDRVGSYCCQCAAGTSGKNCCEVNVNDECHSN 684
Oy 336 -CVNGKCYHVS-----ACLEAGFAGERCEAR-----LCPEGLYGIKCDKR 367
Db 685 PCNNATC--IDGINSYKCOCPVFTGQHCERNVDECISPPCANNGVCIDQVNGYKCE-- 740
Oy 368 CPCHLENTH-----SCHPMSECBACKPMGSLYCNETSPPGYGACQ---QICS-- 414
Db 741 CPROGYDAHCLSDUDECSNPNVCNVEGRCEDEGINETICH--CPPGYTGRCLELIDECSSN 798
Oy 415 -CONGADC-DSVTG-KCTCAPGFKGIDCT-----PCPLG----- 446
Db 799 PCOHGCTCYDKLNFSCQCMPGYTQKCEETNIDCVTPNPGNGGTCIDKVNKCYKCVKP 858
Oy 447 TYGINCSR---CG---KKNVACSP---VDGCTCAGMHGVCDSIRCSBGWFGFCN 496
Db 859 FTGRCCSKMPCANRRKNAKCTPSSNFTDFSCCTKLGTYGRYCDIDIDE-----CS 912
Oy 497 LTCOCLNGACNTLDGT--CTCAPGMWGEKCEL-----PCOOGTGLMCAERCDSDH 546
Db 913 LSSPCRNAGASCLNPGSYRCLCTKGYEGRDCAINDDCASFPQNGR-----TCLD 963
Oy 547 ADGCHPTTGHRCLEPGMSGVHCDVSACAEGRWPN-C-LLPCYCKNGASCSP--DDGICECA 603
Db 964 GIGDY-----SCLCYDGFEDKHCEITDINE-----CLSQP--CONGATCSQYVNSTCTCP 1011
Oy 604 PGFRGTTCOR-----ICSPGFYGHRCSTQCPQCV--HSSGPPCHHITGLCDC 647
Db 1012 LGFSGINCOTNDECTESSCLNGSSCIDGINGNCS-----CLAGYSAGANCOYKLNKCD 1066
Oy 648 LPGETGALCNEV-----CPSGRFRKNC-----GICTCTNNGTCNPIDR--SCOCY 692
Db 1067 NPLCINATGCHBQNNNEYTCHCPSGFTGKQCSFYVMCQSPENAKTSQMKHOFSCKSA 1126
Oy 693 GWIGSDSQCPRAHMKPNCTHTCNCH-----NGAFCSAYDGE--CKGTPG 736
Db 1127 GWTGKGLD-----VOTISCODAADRKGLSLRQLCNNGTCKDYGSHVYCYSOG 1174
Oy 737 WTLGYTORCPLGFYGRKDALICO---CONGADCDHISG--QCTCRFGMRHCEQ--- 787
Db 1175 YASSTCKEID-----ECOSQPCQNGGTCRDILGAVECCORRGFGOGNCELIDD 1224
Oy 788 -----KCPSTGYGYGC--ROICD--CLNNSTG-DHITG--TCY 818
Db 1225 CARNPCONGTCHDRVNVFSCSCPRTGMIICELINKDCKRPGACHNNGSCIDRVGFEV 1284
Oy 819 CSPGMKARGCDQAGVIIVGNLN 840
Db 1285 CQPFVARGCE-----GDIN 1299

```

```

US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitajewski, Jan
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
; FILE REFERENCE: 53863-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/467, 997
; CURRENT FILING DATE: 1999-12-20
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1964
; TYPE: PRT
; ORGANISM: mouse
US-09-467-997-1

```

```

Query Match      14.1%; Score 953.5; DB 4; Length 1964;
Best Local Similarity 25.0%; Pred. No. 2.8e-55;
Matches 297; Conservative 75; Mismatches 303; Indels 513; Gaps 61.

```

```

Oy 109 CADKCVHGRCI---APNTCOCEPGMGNGNSSACDGDHMGHCHSRCKNGALC----- 160
Db 32 CAN---GGTCLRLSRGGGICQCAPFLEGTG-----QFPDPCNDTOLCKNGSGCAL 81
Oy 161 -----NPTTG--ACHCAAGFRGNRCEDR----- 181
Db 82 PTPSSRSPTSLTPHFSCOTCPSPGTDRCDTHBELCPSPSCNGHCYVQASRPQCS 141
Oy 182 CEQGTGYNDCHOR--CO---CONGATC--DHVTGCRCPPGYTGAFCE----- 223
Db 142 CEPMTGQCOLRDFCSANPNANGVCLATYPOIDCRCPPEGTEGTCERDINECFLEBGP 201
Oy 224 -----LCPPKKGPOCEOR---CP---CONGVY-----HVTGEGSC 255
Db 202 CPGTSGHNTLGSVOCLCPVQOEGPOCKLRGACPPSGCLNGTQLYPEGHSTFHLCLC 261
Oy 256 PSGMNGTYCGO-----PCPEGRFRKNCSQ---ECO----- 282
Db 262 PPGFTGLDCENNPDDCVRHOCONGATCLDGLDYTCPCPKTWKMDSESDIDECEARGP 321
Oy 283 -CHNGTCDATG--OCHCSPGYTGERCODE-----CPVG 314
Db 322 RCRNGGTQNTAGSFHCYCVSGMGAGCEENLDDCAATCAPGSTCIDRVGSFSCCLPPG 381
Oy 315 TYGVLC-----AETCO----- 325
Db 382 RTGLCHLEDMCLSPCHVNAQSTNPLTGSTLICIQPGYSGSTCHODLDECOAAGPS 441
Oy 326 -CVNGKCYHVSAG--CLEAGFAGERCEAR-----LCPEG 358
Db 442 PCEHGSSCINTPPGSFNCLCPGYTSRCEADHNECLSQPHGSTCLDLLATFHLCLPPG 501
Oy 359 LYGIKCD---KRC---PCHLENTHSCHPMSG--BCACKPMGSSGLYCN----- 398
Db 502 LBEGLCEVYNECTSNPC--LNOACHDLNLNGFQCLDLPGETGARGCKDDECSSTPCAN 559
Oy 399 -----TCSBGFYGEACQICIS-----CONGADCDSVTGK--CTCAPGFKGI 437
Db 560 GGRCRDQPAFYECBLPFGSPHCEKEVDECLSPPCVVASCLDLPALFCLCRPFTGQ 619
Oy 438 DCSPT-----CPLGTGT-----NCSNRG----- 457
Db 620 LCEVPLCTPNMCOPOGQOCQOEHAPCLCPDGSPGCVAPARDNFCBHGHGORSILCYCDEG 679
Oy 458 -----CKNDVCSPYDG--SCTCKRGMHGVNCS---TICPSGTW----- 491
Db 680 WTPPECTELGCGISTPCAHGATGCHPQSPSGYNTCTCPAGYAGLTCSESEVTAACHSGPCLNG 739
Oy 492 -----GFGCN-----LTCOCLNGACNTLDGT--CTCAPGMWGEK 525
Db 740 SCSTIREGYSCTCLPSHTGHCOTAVDHCVSASCLNGGTCVNRPGTFELCATGAFQGLHC 799

```

```

RESULT 14
US-09-214-278-2
: Sequence 2, Application US/09214278
: Patent No. 6291210
:
: GENERAL INFORMATION:
:
: APPLICANT: Sakano, Seiji
: APPLICANT: Itoh, Akira
:
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
:
: FILE REFERENCE: KP-8576
:
: CURRENT APPLICATION NUMBER: US/09/214,278
:
: CURRENT FILING DATE: 1999-01-26
:
: NUMBER OF SEQ. ID NOS: 32
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ. ID NO. 2
:
: LENGTH: 1055
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-214-278-2

```

```

0Y      278  --SODECO---CHNGOTDAAATG--OCGSCSPYGEROQ---DECPVGTGLVCAETKOCQV 327
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      366  LDAMCBCEKPLCLNAPSKNLIGYTCCITPMHGINCHINIVNOC-----RGQOQ 444
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      328  NGKACVH-VSG-ALICEAGFAGEKCEARLCEPGLYGIKCDKRCPCLEHTHSCHPMSGEC 385
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      445  HGGCTKLDLVNQTQVCCRCRGGRGRICE-----LERKC-ASSPCH-----SGG- 485
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      386  ACKPBMWSGLYCNETCSBPFIYGEACQ---QICS---CONGADCDSDVTGK--CTCAPGFKGI 437
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      486  LCEDLAAGFHCH--CPQGSFSLPLEYVDVLCBPSPCNAGRCVNLBEDDYCACPDGFGK 543
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      438  DCST--PCPLGTIYGINCSSBCCGCKNDA-----VCSFVDSCTCTKAGMHGVDCS 483
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      544  NCSVPRECCPGGA-----CRAVIDGGSDAGMPGTAAAGVCGP--HGRVCSOPGG--NFS 595
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      484  IRCPSGTMGFCN-----LTCQCLNAGC-NTLUDG-TCYCAGMGKECE-----LP- 528
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      596  CIGCSGFGTGTCHENIDDCJLQOPCRNNGTGTIDEVDAFRCPCPBGSGEILDTPNPNDCLPD 655
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      529  -----CODGTGTGLNCAR---CD---CSHADGGHPT--TGHCKCLPG 562
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      656  PCHSRGRCYDLVNDFYACADDMGKGTCHSRERQCAAYTTSNCGTGYDSGDFTFRACPPG 715
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
0Y      563  MSGVHDCSVCAEGRMGPNCSLPCYCKNKAGSC--SPDGICECAPGFGTGTQORICSPGFY 620
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      716  WKG-----STCAVAK-NSSC-LPNCVYNGGTCTVSGASFSCTICXDGWEGRTCT----- 761
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

RESULT 15
US-09-214-278-3
: Sequence 3, Application US/09214278
: Patent No. 6291210
: GENERAL INFORMATION:
: APPLICANT: Sakano, Sei11
: APPLICANT: Itoh, Akira
: TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
: FILE REFERENCE: KP-8576
: CURRENT APPLICATION NUMBER: US/09/214,278
: CURRENT FILING DATE: 1999-01-26
: NUMBER OF SEQ ID NOS: 32
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 1212
: TYPE: prt
: ORGANISM: Homo sapiens
US-09-214-278-3

```

Best Local Similarity: 26.8%; Pred. No. 1,4e-44;
Matches 309; Conservative 93; Mismatches 417; Indels 332; Gaps 74;

```

OY 15 LCHWMTGAPLNEEDNVCNWSHESVTVQESVPHFDDIYITSCDILIMFECTHRY 74
DB 135 LLERYSHAMINPEDMKSLHSGVAHLELOIRVNCDEYISATON-----KECPRN 189
OY 75 SYRTAVRHGEKTYWRKRSOCCPGFYESGEMCVPH-CADKC--VHGRIAPNTCCCEBGMG 131
DB 190 DF---FGHYTCDOYGKNA-CMDGW--WGRCKEAVCKOGNLLHGCTVGECCSGWQ 243
OY 132 GTNCSSACDDHMGPHCTSCCKNGALCNPTIGACCAAGFRGKREDCEOGCTYNDG 191
DB 244 GRECD-----ECVPRGCVHSGCVPR--WOCNCEITNMGILLCDKDL-----NYC 285
OY 192 HORCOCONGATCDHVTE---CROPPGYTAFCEEDLCPGKHGPOCEORC---PCONGV 245
DB 286 GSHHPCITNGTICINAEPOYRCTCPDGYSGRNCB-----KAEHACTSNPCANGGS 335
OY 246 CHHVTG--ECSRPSGMWMTVC-----GQP-----CPEGRFGKNC- 277
DB 336 CHEVPSGFECHCPSGMGTFCALDIDECASNPCAAGGTCVDYDGFECICPEGWVATCQ 395
OY 278 --SQEQO---CHNGTCDATG--QCNCSPGYGERCQ---DECPVGTGYLCAETCQCV 327
DB 396 LDANECEGKPLNAFSCKNLIGGYCDIPGMKGINCHINVNDG-----RGOQ 444
OY 328 NGKCYH-VSG-ACICEAGFAGERCEARLCPGLYGIKCDKRCPEHLENTHSCHPMSGEC 385
DB 445 HGCTCNDLVNGYOCVCRGGRHCE-----LERDKC-ASSPCH-----SGG- 485
OY 386 ACKPMSGLYCNETCSBGFYEACQ---QICS---CONGADCSVTGK--CTCAPGFKGI 437
DB 486 LCEDLDAGFHH--CPQGFSGPLCEVDVLDCEBSPCRNGARCYNLEGDYCACPDGFGK 543
OY 438 DCST---PCPLGTGYTIGCSRCCKKND-----VCSPIVDSCTCKAGHGVDCS 483
DB 544 NCSVPREPCPGA---CRVIDCGSDAGPMGTAAAGVCGP-HGRCVSQPG--NFS 595
OY 484 IRCPSGTWFPGCN-----TTCOLINGAC-NTLDG--TCTCAPGMGEKCE-----LP- 528
DB 596 CIDDSGFTGYCHENTIDDLGQPCORNGTICIDEVDARCFPCPSWEGELCDTNPNDCLPD 655
OY 529 -----CODGTYGLNCAER---CD---CSHADGCHPT--TGHCRCLPG 562
DB 656 PCHSRRCYDLYNDFYCACDDGMKMGKTHSREPOCAVATCSNGTCTYDSGDTFRCACPG 715
OY 563 WSGVHDSVCAEGRMWGNCLPCTCKNGASC--SPDDGICECAPGFRGTTCQRLCSBGFY 620
DB 716 WKG---STCAVAK-NSSC-LPNPCVNGTGVSGASFSICIRDGWEGRTCT----- 761
OY 621 GHRCSOTPCQVSHSGFCHHITGL---CDLPGFTGALCN---EVCPSGRFGKNCAGIC 673
DB 762 -HNTNDCNPLPCYNGGIC--VDGVNMFRCBAGFAGPDCRINIDECQS-----SPCAYGA 814
OY 674 TCTN--NGTCNPIRSCQCPYGIISDC-----SQPCPPAHMGPNCIHTCNCHN 720
DB 815 TCVEIDING-----YRSCPPGRAGPRCQEVIGFGRSCMSRGTPFPH-GSSWEDCNS-- 865
OY 721 GAFCSAYDECKCTPMTGLYCTGRCPGLGYGDCALICQONGADC-DHISGOC---TC 776
DB 866 ---CRCLDGRDCKYWCQ--WKPCLLA--GQPEALSACPLGQRLERAPGOCLARPC 917
OY 777 RT-GFMGRHCEQKCP-----SGTYGCGORICDCLNNSTCDHI---TGTCYS-----PG 822
DB 918 EAMGEGGAEPSTPCLPRSGHLDNNCARLTLFNR---DHVPOGTTVGAICSGINSIPA 974
OY 823 WKGARCDOAGVIIVGNLNS-LSRTSTA-----LPADSYQIGAINGIIILVVLVLEL 873
DB 975 TRAVARDRLVLLCDRASSGASAVEAVSFPARDLPDSSLIGGAHAIVAAI----- 1027
OY 874 ALFTIIRHKOKKSSNPVATYTPAMRVVNADYTTISGTLPHSNGNANSHYFTNPSTYHTL 933

```

```

DB 1028 -----TORGNSSLLAATVEVKVETVVTGSGSTGLLVPVLCGA-----FSYLMACY 1073
OY 934 TOCA--TSPHYNNRDMRTVTK--SKNQ-----LFVNLKNVNFGR 970
DB 1074 VLYWMTTRKRKRERERSRLPRESANQMAPLPIRNPTRPGHMDVLYQCKNFTPPR 1133
OY 971 -----GPVG 974
DB 1134 RADEALPGPAG 1144

```

Search completed: May 9, 2003, 11:59:20
Job time: 58.4565 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 07:56:52 ; Search time 60.7648 Seconds

(without alignments)
2499.898 Million cell updates/sec

Title: US-10-092-390-2

Perfect score: 6744
Sequence: 1 MVSLNSGSLFLLCHWT.....SSPKQDSGSSSSSSSSSE 1140

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

A.GeneSeq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3107.5	46.1	878	22	ABG08033	Novel human diagno
2	2506.5	37.2	1050	22	AA866267	Human TANGO 272 SE
3	1909	28.3	636	22	AA866269	Rat TANGO 272 SEQ
4	1522	22.6	384	22	AA875479	Human colon cancer
5	1466	21.7	321	22	ABG27639	Novel human diagno
6	1272	18.9	466	22	ABG22559	Novel human diagno
7	1252	18.6	434	22	AB866756	Drosophila melanog
8	1241.5	18.4	762	22	ABG08032	Novel human diagno
9	1192	17.7	474	22	AA72715	HFTU08 clone huma
10	1169	17.3	269	22	ABG08031	Novel human diagno

11	1034.5	15.3	2444	23	AB87821	Constitutively act
12	1014.5	15.0	1473	23	AAE18208	Human Molib protei
13	1012.5	15.0	2471	20	AA706816	Human Notch2 (huma
14	1011	15.0	241	22	AA72712	HttLH4 clone huma
15	1002.5	14.9	587	22	ABG03826	Novel human diagno
16	980	14.5	2469	23	AAE18207	Human Molia protei
17	978.5	14.5	2703	22	AB860266	Drosophila melanog
18	974	14.4	2321	19	AAW49658	Human Notch3 prote
19	961.5	14.3	447	22	AB866268	Human TANGO 272 S
20	953.5	14.1	1964	20	AAW95557	Mus musculus notch
21	940	13.9	1872	19	AAW68510	Partial human Notc
22	919	13.6	4618	22	AAW39043	Human polypeptide
23	895	13.3	209	22	ABG27638	Novel human diagno
24	813	12.1	833	21	AA85616	Murine nurse cell r
25	813	12.1	833	22	AA860392	Murine nurse cell r
26	808	12.0	865	23	AAW47668	Mouse nurse cell r
27	808	12.0	866	22	AAW60393	MOLab protein sequ
28	808	12.0	866	22	AAW60394	Human nurse cell r
29	802.5	11.9	831	23	AB880609	Human sbg115B5RRC
30	800	11.9	871	22	AAW60395	Human nurse cell r
31	800	11.9	871	23	AB880610	Human sbg115B5RRC
32	789	11.7	1238	23	ABW07823	Human notch agonis
33	788	11.7	1212	19	AAW44299	Human serrate 2.
34	787	11.7	1055	19	AAW44298	Human serrate 2 pr
35	785	11.6	129	22	ABG22555	Novel human diagno
36	785	11.6	129	22	ABG27637	Novel human diagno
37	783.5	11.6	884	23	AAW47667	MOLab protein sequ
38	781	11.6	1257	17	AAW05834	Human Serrate-2 (H
39	781	11.6	1257	21	AAW59598	Human Serrate prot
40	779	11.6	296	22	AAW56069	Skin cell protein,
41	779	11.6	296	23	ABW72269	Murine EGF family
42	779	11.6	299	21	AA76014	Murine EGF family
43	779	11.6	299	21	AA76017	Skin cell protein,
44	779	11.6	299	22	AAW59593	Skin cell protein,
45	779	11.6	299	22	AAW56016	Skin cell protein,

ALIGNMENTS

RESULT 1
ABG08033
ID ABG08033 standard; Protein: 878 AA.
XX
AC ABG08033:
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #8024.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW Food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YF:
XX WPI; 2001-639362/73.
XX N-PSDB; AAS72220.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT

PT responsible for genetic disorders or other traits and to assess
 XX biodiversity -
 PS Claim 20; SEQ ID No 38392; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantifying a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 878 AA:

Query Match 46.1%; Score 3107.5; DB 22; Length 878;

Best Local Similarity 59.5%; Pred. No. 4.4e-155;

Matches 594; Conservative 15; Mismatches 69; Indels 321; Gaps 13;

QY 338 ACICEAGFAGEGCEARLCEPEGLYGIKDCRCPCHLEN--THSCHPMSEGCACKPGMSGLY 395
 DB 5 ALICQLTV-----C-----ISAQLICPFAMEQQLVACCHPMSEGCACKPGMSGLY 50
 QY 396 CNETSPGTYGACQOICSCQNGADSDSYTGKTCAPGFKGIDCSTPCPLGTGYNCSR 455
 DB 51 CNETSPGTYGACQOICSCQNGADSDSYTGKTCAPGFKGIDCSTPCPLGTGYNCSR 110
 QY 456 GCCKNDVAVSPYDGSCTCAGHGVDCSTRCPSTGTFPCNLTCCCLNGACNTLDGCT 515
 DB 111 GCCKNDVAVSPYDGSCTCAGHGVDCSTRCPSTGTFPCNLTCCCLNGACNTLDGCT 170
 QY 516 CAPGNRGEKCELPQDGYTGLNCAERDCSHADGCHPTGHCRCLEPGMSGVHCDVCAEG 575
 DB 171 CAPGNRGEKCELPQDGYTGLNCAERDCSHADGCHPTGHCRCLEPGMSGVHCDVCAEG 230
 QY 576 RRGPNCSLPCYCKNAGSCSPDDGICECAFEGTTCQRLCSGFTGHRCSQTCPOCVHSS 635
 DB 231 RRGPNCSLPCYCKNAGSCSPDDGICECAFEGTTCQRLCSGFTGHRCSQTCPOCVHSS 290
 QY 636 GPCHHITGCDCLPGFTGLNCEVPSGRFGKNCAGICTNNGCNPIDRSCQCCPGMI 695
 DB 291 GPCHHITGCDCLPGFTGLNCEVPSGRFGKNCAGICTNNGCNPIDRSCQCCPGMI 332
 QY 696 GSDSCQPCPP-----AHMGPNCIH-----TCNCHNGAFCSAYDECKTSP-- 735
 DB 333 PAPTEPEYTPPLDVLVSYFVGLAVTALNHNHLP.TGAGTASTRATATIM-----ELSAAPIM 387
 QY 736 -----GWTGL-----YCTORCLPGTYGKDCALICQCONGADCDHISGQCTCRG 779
 DB 388 GNVNALLAGQGSTALRRSPRHSCRAAASPWFYKDCALICQCONGADCDHISGQCTCRG 447
 QY 780 FMGRHCEOK----- 788
 DB 448 FMGRHCEOKVRRPMDHRLMTALGGGGVTTTMMKTEFKSILFFWALPSSSYFWNVAOS 507
 QY 789 ----- 788
 DB 508 LKRSSAFPMALAEPSGSHIGQYIRMGGLVAGOSLLPCAIVTVYSATIMIGMLSSST 567

QY 789 -----CPSGTGYCCRCIO 802
 DB 568 ILGVQVSLNRNP.LKGLSSNAGLAVRDSLAPNSQGMKATFDEPSPLECPSPGTGYCCRCIO 627
 QY 803 DCLNNSTCDHITGTCTCSPGMRGARCDAQVITVGNLSLSTSTALPADSDQIQAINGI 862
 DB 628 DCLNNSTCDHITGTCTCSPGMRGARCDAQVITVGNLSLSTSTALPADSDQIQAINGI 687
 QY 863 ILLVLYFLALFFIIRHKQKKESSMPAYTTPAMRVNADYITSGTLPHSNGNANS 922
 DB 688 ILLVLYFLALFFIIRHKQKKESSMPAYTTPAMRVNADYITSGTLPHSNGNANS 747
 QY 923 HFTNPSTHTLTOCATSP-HVNNRDMRYTCKSNQOLFVNLKNVPGARKGPVDCGTGLP 981
 DB 748 HFTNPSTHTLTOCATSPERSSTGDMRYVHEKK----- 781
 QY 982 ADMKRGVINEGALGDRSYGKSLKDLGKSEYNSSNCSLSSENPRATIKDPVLP 1041
 DB 782 ----- 781
 QY 1042 KSECEGYEMKSPARRDSPAIEINNSTSANRVYEVEPTVSYGVFSNNGRLSODPYDL 1101
 DB 782 QSTVC--ESMKSPARRDSPAIEINNSTSANRVYEVEPTVSYGVFSNNGRLSODPYDL 839
 QY 1102 PKNSHIPCHYDILLPYRDSSSSPKQEDSGGSSSSSSSE 1140
 DB 840 PKNSHIPCHYDILLPYRDSSSSPKQEDSGGSSSSSSSSSE 878

RESULT 2

AA066267 standard; Protein; 1050 AA.

AA066267;

05-APR-2001 (first entry)

Human TANGO 272 SEQ ID NO: 14.

Membrane associated protein; secreted protein; human; mouse; rat;
 INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295;
 TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
 haematopoietic disorder; neural disorder; hepatic disorder;
 neoplastic disease.

Homo sapiens.

WO200100673-A1.

04-JAN-2001.

29-JUN-2000; 2000WO-US18198.

30-JUN-1999; 99US-0345464.

(MILL-) MILLENNIUM PHARM INC.

Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;

WPT: 2001-050128/06.

N-PSDB; AAF27787.

Isolated secreted or transmembrane proteins are used for diagnosis and
 treatment of neoplastic and haematopoietic disorders e.g. T cell
 disorders, cancer and tumours -

Claim 9; Page 22-229; 294pp; English.

The present invention provides the protein and coding sequences for a
 number of membrane associated and secreted proteins from human, mouse
 and rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
 TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
 involved in signal transduction and the sequences can be used in the

CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders.

XX Sequence 1050 AA;

Query Match 37.2%; Score 2506.5; DB 22; Length 1050;
Best Local Similarity 40.5%; Pred. No. 1.6e-123;
Matches 490; Conservative 111; Mismatches 345; Indels 263; Gaps 30;

```

QY 14 LLLCHWIGTASPLLEDPNVCSSHEISVYQSEYPRHPDQIYTTCTDILNM----- 66
DB 9 LLLAVGLRAGTLNPSBPNTCSFEWSEFTTTTKESHSPFSLPSEPE--RPWEGPHTCP 66
QY 67 -----FKCTHRVSYR-----TAY 80
DB 67 SPOTOKILLASRDSFCVACVAGVQVMDRSLADPQTGNALSMRQPRVYLSCAPLASBQH 126
QY 81 RHGEKTYRRKSOCCPGFESEGEKCVPHADKCVHGRCIAPNTQCEPFGMGTCSSA-- 138
DB 127 TVVYKTDHRLQCCCHGFEGRGFCVPLCAQECVHGRCVAPNQCQVPMGRWDDCCSAPN 186
QY 139 -----CDGDHMGPHCTSRQCKNGALCNPTGACHCAAGFGMRCEDRCQGTGNDCHQR 194
DB 187 CLDPTGTYGPACQFRQCC--HGAPCDPQTGACFCPAERTGSPCDVSCQGT----- 237
QY 195 CQCCNGATCDHVTGECNCPPTGTGAFCEDLCPRGKHGRQCRQPCONGVCHVTEGCS 254
DB 238 -----SSGFPC-----PSTH-----PCQNGVYQTPQSSCS 262
QY 255 CPSSMGTVCGQPCPEGRFGKNCQEQCHNGGTCDATGQCHGCSPGYTGRCODECPVG 314
DB 263 CPPEMNGTICSLPCEGFHGNCSQEQCRCHNGICLDRFTGQCRAPGTGBRCHECPVG 322
QY 315 TYGVLCALCTCCVNGKCYHVSACLBAGFAGRCENARLCBELYXGICDKRCPCHELEN 374
DB 323 RFGDCAETCCADACFPANGACLCENHFTGDRCTDRLCDPDEFGYGLSCAPCTCIREH 382
QY 375 THSCHPMSEGCACKPMGSGLTGNETSPGFGACQOQICSCONADDDSVYKCTCARGF 434
DB 383 SLSCHPANGECSCLPGMAGLCHNESCPDQTHGRCQCHLCLHGGVCAVATSGILQCAAGY 442
QY 435 KGIDCSTPCPLGTGYNCSRCGCKNDAYCSPVDSCTCKAGWGVDCSIRCPSGTWFG 494
DB 443 TGPHCASLCPEDTGVNCSARCSENAIACSPIDGECVCKGMRGNKSVPCPTGTFGS 502
QY 495 CNLTQCLNGACNTLDTGTCAFGMGEKCELPQDGTGVLNACERDCSHADCHPTT 554
DB 503 CNASCQCAHEAVYCSPTGACTCTPGMHGANCOLPCPKQGFEGECASRDCDHSQCDPVH 562
QY 555 GHCRCLPGMSGVHCDVCAEGRMGNCSLPYCKNGASCSPDDGICGCAPEFRGTGORT 614
DB 563 GRCOCQAGMGMGARCHLSCPEGLMGNCSNTCTCKNGGCTCPENGCYCAPFRGSPSCRS 622
QY 615 CSPFGYHRCSTQCPQCVHSSGPHNITGLDCLPFGTGALENEVCSGRRGKNCAGICT 674
DB 623 CQPERYGR-----CYP-----CK 636
QY 675 CTNNGTCNPIDRSCQCTFGMTGSDSCQPCRAHWPNCITCNCHNGAFCSAYDGECT 734
DB 637 CANHSFHPNSNGTCYCLAGMTGTPDCSQPCRGHNGENCAQCCCHGGCTCHPDGSCICP 696
QY 735 PGWGLVCTONCRPLGFGKDCALICOCQNGADCDHISGQCCRGFGMRHCEQKCPSTY 794
DB 697 LGWTHGHLCECPGRTGANCSPQCCGPGKEK-----HPE----- 732
QY 795 GYGCROICDCLNNTCHITGTCTCSPGMKARGCDQAGVITVGNLNSLSRSTALPAD-- 852
DB 733 -----TGACVCPGHSAPCR-----IG-----IQEFTYMPPTPV 763
QY 853 SY-QIGAIAGITLIVLVVFLALFIIRKOKGKESMPAVTTPAMRVVNADYTSIGT 911
DB 764 AYSIGAVTGAIVAGSLVALVALFVIGRHMOKGKEHHHLHVAVSSG--RLDGSSEVMPDV 822

```

```

QY 912 LPHSNGNANSHYFTNPSTYHTLQCATSPHYNNRDRMTVTKSKNOLFVNKN--VNPGRK 970
DB 823 PP-----SYSHYSNPSYHTLSQCSPPNPPNK-----VGPPLFASLQNERG-- 866
QY 971 GPVG-DCTGTLPADMKH-----GGYLNELAFGLDRSYMKSL-----KDLGNSEY 1016
DB 867 GAQCHDNHTLPLADMKHRRPEPPPLDR--GSSRLDRSYSYSGNPGPFYDKGLISEEL 925
QY 1017 NSSMCSLSSSNPYATTKDPPVLLPKSSECYVEKSPARDSPYAEIINSTSANRNYE 1076
DB 926 GASVASTL--SSENPYATLRDPLSLGCPRESSYMEKGPSSGAPRQPPQFWDQRR-- 981
QY 1077 VEPYTVYQGVFSNNGRLL-----SODP-----YDLPKNSHIPCHYDLPYVDS 1119
DB 982 -QPPQDRDSGTIEQPSPLIHDRDSVGSOPPLPPLPCHGHTSPKNSHIPGHYDLPVYRHP 1040
QY 1120 SSSP-KQED 1127
DB 1041 PSPPLRRQD 1049

```

RESULT 3

AAB66269 ID AAB66269 standard; Protein: 636 AA.

AC AAB66269;
DT 05-APR-2001 (first entry)
XX Rat TANGO 272 SEQ ID NO: 20.

DE Membrane associated protein; secreted protein; human; mouse; rat;
KW INTERCEPT 340; MANGO 003; MANGO 347; TANGO 272; TANGO 295; TANGO 354;
KW TANGO 378; skeletal disorder; cardiovascular disorder; renal disorder;
KW haematopoietic disorder; neural disorder; hepatic disorder;
KW neoplastic disease.

OS Rattus sp.

PN W0200100673-A1.

PD 04-JAN-2001.

PF 29-JUN-2000; 2000MO-U018198.

PR 30-JUN-1999; 9905-0345464.

PA (MILL-) MILLENNIUM PHARM INC.

PI Barnes TM, Fraser CC, Wrighton N, Myers P, Busfield SJ, Sharp JD;

DR WPI; 2001-050128/06.

DR N-PSDB; AAF27791.

PT Isolated secreted or transmembrane proteins are used for diagnosis and
PT treatment of neoplastic and haematopoietic disorders e.g. T cell

PT disorders, cancer and tumours -

PS Claim 9; Page 238-240; 294pp; English.

CC The present invention provides the protein and coding sequences for a
CC number of membrane associated and secreted proteins from human, mouse and
CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
CC involved in signal transduction and the sequences can be used in the
CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders.

XX Sequence 636 AA;

Query Match 28.3%; Score 1909; DB 22; Length 636;
Best Local Similarity 45.1%; Pred. No. 2e-92;
Matches 328; Conservative 77; Mismatches 212; Indels 110; Gaps 9;


```
XX 18-FEB-2002 (first entry)
DT
XX
DE Novel human diagnostic protein #27630.
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS91826.
XX
XX Claim 20; SEQ ID No 57998; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 321 AA;
XX
XX Query Match 21.7%; Score 1466; DB 22; Length 321;
XX Best Local Similarity 90.3%; Pred. No. 1.7e-69;
XX Matches 241; Conservative 0; Mismatches 0; Indels 26; Gaps 2;
XX
XX 137 SACCDDHMGPHCTSRCCCKNGALCNPITGACHCAGFGMRCEDECGTGNDCHORCO 196
XX |
XX 9 SACCDDHMGPHCTSRCCCKNGALCNPITGACHCAGFGMRCEDECGTGNDCHORCO 68
XX
XX 197 CONGATCDHYTGECRCRPGYTGAFCEDLCPRKHGRPOCEORPCONGGVCHHVTGECSCP 256
XX |
XX 69 CONGATCDHYTGECRCRPGYTGAFCEDLCPRKHGRPOCEORPCONGGVCHHVTGECSCP 128
XX
XX 257 SGM-----GTYCGQPCPEGRFGKNCSDCCOCHNGSTCDATGQCHC 298
XX |
XX 129 SGMMLSPGWRPIKFSKSLXMGTVCGQPCPEGRFGKNCSDCCOCHNGSTCDATGQCHC 188
XX
```

```
OY 299 SPGYTGER-----CODECPVGYGVLCATCCQVNGKCYHVSAGCLCEAGPAGRC 350
XX |
DB 189 SPGYTGERAAVPDVRKCODECPVGYGVLCATCCQVNGKCYHVSAGCLCEAGPAGRC 248
XX
OY 351 EARLCEGLYGIKCDKRCPCHELENTHS 377
XX |
DB 249 EARLCEGLYGIKCDKRCPCHELENTHS 275
XX
RESULT 6
ABG22559
ID ABG22559 standard; Protein: 466 AA.
XX
XX ABG22559;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #22550.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX N-PSDB; AAS86746.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID No 52918; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 466 AA;
XX
XX Query Match 18.9%; Score 1272; DB 22; Length 466;
XX Best Local Similarity 76.9%; Pred. No. 3.5e-59;
```

	Matches	220: Conservative	3: Mismatches	23: Indels	40: Gaps	5:
QY	118	CIAPPTCCCEBFWMGINTSSACDDHMGPHCTSRQCKANGALCNPTGACHC-AAAGFGW	176			
Db	39	CFSP---OSQPMVITGVTPAPAG-----ASAKKGLCATPTTGTTCCLMLRGFRW	85			
QY	177	RCEDECEQGTGANDCHQKCCONGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPOCEQ	236			
Db	86	RCEDECEQGTGANDCHQKCCONGATCDHVTGECRCPPGYTGAFCEDLCPGKHGPOCEQ	145			
QY	237	RCPCONGVCCHVHTGECSCPSGWM-----GTVCGQPCPEGRRGKNS	278			
Db	146	RCPCONGVCCHVHTGECSCPSGWMSPFGWRPIXFSKSLXMGGTGVCQPCPEGRRGKNS	205			
QY	279	QECOCNHNCGTDDAATGQCHCSPGYTGER-----CODECPVGTGYVCAETCCQVNG	330			
Db	206	QECOCNHNCGTDDAATGQCHCSPGYTGERAAVPDVKRCODECPVGTGYVCAETCCQVNG	265			
QY	331	KCYHVSACILCEAGPAGERCEARLCPGLGYGIKCDKRCPCHELNTH	376			
Db	266	KCYHVSACILCEAGPAGERCEARLCPGLGYGIKCDKRCPCHELNTH	311			
RESULT 7						
ABB6756						
ID	ABB6756	standard; Protein: 434 AA.				
AC	ABB6756;					
XX						
DT	26-MAR-2002	(first entry)				
DE	Drosophila melanogaster polypeptide SEQ ID NO 27060.					
KM	Drosophila; developmental biology; cell signalling; insecticide;					
KM	pharmaceutical.					
XX						
OS	Drosophila melanogaster.					
XX						
PN	WO200171042-A2.					
PD	27-SEP-2001.					
XX						
PE	23-MAR-2001; 2001WO-US09231.					
PR	23-MAR-2000; 2000US-191637P.					
PR	11-JUL-2000; 2000US-0614150.					
PA	(PEKE) PE CORP NY.					
XX						
PI	Venter JC, Adams M, Li PWD, Myers EW;					
DR	WPI; 2001-656860/75.					
DR	N-PSDB; ABL10859.					
PT	New isolated nucleic acid detection reagent for detecting 1000 or more					
PT	genes from Drosophila and for elucidating cell signalling and cell-cell					
PT	interactions -					
PS	Disclosure; SEQ ID NO 27060; 21pp + Sequence Listing; English.					
XX						
CC	The invention relates to an isolated nucleic acid detection reagent					
CC	capable of detecting 1000 or more genes from Drosophila. The invention is					
CC	useful in developmental biology and in elucidating cell signalling and					
CC	cell-cell interactions in higher eukaryotes for the development of					
CC	insecticides, therapeutics and pharmaceutical drugs. The invention					
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA					
CC	sequences (ABL01840-ABL16175) and the encoded proteins					
CC	(ABBS7737-ABBS72072).					
CC	The sequence data for this patent did not form part of the printed					
CC	specification, but was obtained in electronic format directly from WIPO					
CC	at ftp.wipo.int/pub/published_pct_sequences.					
XX						
XX	Sequence 434 AA:					

	Query Match	18.6%	Score 1252:	DB 22:	Length 434;
	Best Local Similarity	45.6%	Pred. No.3.6e-58*		
	Matches 190;	Conservative 55;	Mismatches 170;	Indels 2;	Gaps 1;
Qy	151	RCQCKNGALCNPIITGACHCAAGFRGRCEDRCEQGTGYNDCSHORCOCONGATCDHYTSEC	210		
Db	2	QCCDCLNNVAVCEPFGSGDCBCKAKGYTGARCADICPEGFPGANSEKRCENGGKCHHVSGEC	61		
Qy	211	RCPPRYTGAFCEDLCPFGKHGPOCEQRPCQNGVYCHVHTGECSCPSGWMGTVCQPPDE	270		
Db	62	QCAPFTPLPLCDMRCPRDKHGAQCCQDDPCQDNGKCOPEYGACMKCNPMWTGIVYCNKCPV	121		
Qy	271	GRFKNCSCOECCCHNGGTCDAAATGQCHSPGTYGERCODECPVGYTGYVLAEYTCQCVNG	330		
Db	122	GSYGGCCDESECECYTGACACCHHTTGGCEBPFGYRGRCITDECOLANTYGNCSMTCCANDA	181		
Qy	331	KCYHVSACLCLEAGEAGERCEARLCPESLYGICDKRCPCHELENTSHCHPMSEGCACRPG	390		
Db	182	MCDRANGTCICNPMPTGTAKCAERICAEANKYGDICNRCECDMEHNDLCHPEYGNQCSIG	241		
Qy	391	MSGLYCNETCSPGFYGEACQDLCSCQNGADDSVYGTGCTACRPFYFGIDCSYPCPLGYGI	450		
Db	242	WSSACCTPRCTFRLARYGPNCCELTCCNCKNAKSPVNVGTCLCAPGWRPTEESCESEGTPEQ	301		
Qy	451	NCSRCGCKNDPAVCSPLYVGSCTCKRAGMHGVDCSIRCPSTWFGCGNLTCCQCLNGGACNTL	510		
Db	302	DCALRKCDQNGAKACEPERFGQCLCTAGMKNNIKCDRPCDLNHFGQDCAKYCDCHNNAACRPG	361		
Qy	511	DGTCTCAGWNGEKCCLPCQDGTGYNLCAERDCD--SHADSCPTTGHCRCLPGNSG	565		
Db	362	NGSCTCAGMTGERCKCDTGKFGHDCAKQCCDFNNSLACDATINGRCVCKQDWGG	418		
RESULT 8					
ABG08032					
XX	ABG08032	standard; Protein; 762 AA.			
AC	ABG08032;				
XX					
DT	13-FEB-2002	(first entry)			
XX					
DE	Novel human diagnostic protein #8023.				
XX					
KM	Human: chromosome mapping; gene mapping; gene therapy; forensic;				
KM	food supplement; medical imaging; diagnostic; genetic disorder.				
XX					
OS	Homo sapiens.				
XX					
XX	MO200175067-A2.				
XX					
PD	11-OCT-2001.				
XX					
FE	30-MAR-2001; 2001WO-US08631.				
XX					
PR	31-MAR-2000; 2000US-0540217.				
PR	23-AUG-2000; 2000US-0649167.				
XX					
PA	(HYSE-) HYSEQ INC.				
XX					
PI	Dymanac RT, Liu C, Tang YT;				
XX					
XX	WPI: 2001-639362/73.				
DR	N-PSDB: AAS72219.				
XX					
PT	New isolated polynucleotide and encoded polypeptides, useful in				
PT	diagnostics, forensics, gene mapping, identification of mutations				
PT	responsible for genetic disorders or other traits and to assess				
XX	bioidiversity				
PS	Claim 20; SEQ ID No 38391; 103pp: English.				
XX					
CC	The invention relates to isolated polynucleotide (I) and				

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 762 AA:

Query Match 18.4%; Score 1241.5; DB 22; Length 762;
 Best Local Similarity 44.0%; Pred. No. 2,2e-57;
 Matches 267; Conservative 16; Mismatches 81; Indels 243; Gaps 17;

OY 545 SHADGCHPTGHCRCIPGSHVCHDSVCAEGRWGPNCSIPCYCKNAGSCPDGICECAP 604
 DB 14 SHSVISGAAAGEC-----GGVHCHDSVCAEGRWGPNCSIPCYCKNAGSCPDGICECAP 67
 OY 605 GRRGTTTCORICSPGFYGHRCISQTCPCVHSSGRCCHITGLCDLCEFGALCNEV----C 660
 DB 68 GRRGTTTCORICSPGFYGHRCISQTCPCVHSSGRCCHITGLCDLCEFGALCNEVSSQC 127
 OY 661 PSGRGKNCAGICTCTNNCTGNCNPNIDRSCCQYPGWIGSDSCQPCPAHNPNCIH-TCNCH 719
 DB 128 PSGRGKNCAGICTCTNNCTGNCNPNIDRSCCQYPGWIGSDSCQPCPAHNPNCIH-TCNCH 183
 OY 720 NCAPFSADVGECKTPGWTGLYCTGRCPLGFGYKDCALICQNGADCDHISGQCTCRFG 779
 DB 184 N-----TCQCEPGW-----GGTNCSSVKKQSP----- 205
 OY 780 FNGRHCEOKCPSGTYGYGROICDCLNNTGCHITGTCSPGWKRGARCDQAGVILVGNL 839
 DB 206 -----VCESK-----CEP-WERRP-----GGL 223
 OY 840 NSLSRTSTALPADSYOIGALIGIILVLVFLFLALFIYRHKOKGESSMPAVTYTPAM 899
 DB 224 H----- 224
 OY 900 RYVNAVDTYISGTLPHSNNGNANSHYFTNPSTHTLQCATSPHYNNDRMTVTKSKNNQLF 959
 DB 225 -----WDIAGX-----KHGGLNELGAFGLDRSYMGSLDLGKNSSEYNS 230
 OY 960 VNLKVNPKRGKRGVGDCTGLPADMKHGGYLNELGAFGLDRSYMGSLDLGKNSSEYNS 1019
 DB 231 -----P-----KHGGLNELGAFGLDRSYMGSLDLGKNSSEYNS 267
 OY 1020 NCSLSSSENPYATIKDPVLIPKSE---CG---YVEMKSPARBDSPYAEINNSTSANRN 1073
 DB 268 NCSLSSSENPYATIKDPVLIPKSE---CG---YVEMKSPARBDSPYAEINNSTSANRN 327
 OY 1074 VTEVEPYVSVGVGFSNNGRLSDPPYDLFKNSHIPCHYDLFVRDSSSPKQEDSGGSSS 1133
 DB 328 YSECCPR-SIQQXMASFPESIXP-----PKNSHIPCHYDLFVRDSSSPKQEDSGGSSS 381
 OY 1134 NSSSSSE 1140
 DB 382 NSSSSSE 388

RESULT 9

AAV72715
 ID AAV72715 standard; Protein; 474 AA.
 XX
 AC AAV72715;
 XX
 DT 31-MAY-2001 (first entry)
 XX
 DE HFCU08 clone human attractin-like protein.
 XX
 KW Human; attractin-like protein; HFCU08 clone; food additive;
 KW food preservative; gene therapy; cancer; neuroprotective; cytostatic;
 KW immune disorder; Addison's disease; allergy; cerebroprotective;
 KW autoimmune haemolytic anaemia; autoimmune thyroiditis; nocturnal;
 KW diabetes mellitus; Crohn's disease; multiple sclerosis; cardiac;
 KW rheumatoid arthritis; ulcerative colitis; cardiovascular disorder;
 KW myocardial ischaemia; congestive heart failure; pericarditis; viral;
 KW wound healing; neurological disease; cerebral anoxia; antibacterial;
 KW Alzheimer's disease; communicative disorder; epilepsy; ophthalmological;
 KW infectious disease; immunosuppressive; antiproliferative; vasotropic;
 KW fungicide; chromosome 1p36.3.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Region
 FT Location/Qualifiers
 FT 14..20
 FT /note= "Immunogenic epitope"
 FT 37..52
 FT /label= EGF-like_domain
 FT 55..66
 FT /note= "Immunogenic epitope"
 FT 80..95
 FT /label= EGF-like_domain
 FT 86..92
 FT /note= "Immunogenic epitope"
 FT 123..138
 FT /label= EGF-like_domain
 FT 152..160
 FT /note= "Immunogenic epitope"
 FT 167..182
 FT /label= EGF-like_domain
 FT 176..181
 FT /note= "Immunogenic epitope"
 FT 210..225
 FT /label= EGF-like_domain
 FT 262..273
 FT /note= "Immunogenic epitope"
 FT 279..287
 FT /note= "Immunogenic epitope"
 FT 296..311
 FT /label= EGF-like_domain
 FT 325..332
 FT /note= "Immunogenic epitope"
 FT 339..354
 FT /label= EGF-like_domain
 FT 382..397
 FT /label= EGF-like_domain
 FT 392..397
 FT /note= "Immunogenic epitope"
 FT 404..414
 FT /note= "Immunogenic epitope"
 FT 429..434
 FT /note= "Immunogenic epitope"
 FT 461..474
 FT /note= "Immunogenic epitope"
 XX
 PD WO200116156-A1.
 XX
 PD 08-MAR-2001.
 XX
 PF 29-AUG-2000; 2000WO-US23663.
 PF
 XX 30-AUG-1999; 99US-0151348.
 PR
 XX


```

Db 956 RNCAMCTCDVSYCTCTCPAGFSGIHCENNTPDCTESSCFNGATC--VDGINSFCLCPGP 1013
Oy 694 WIGSDC-----SOP-----CPAHNGPNC---IHNCT---CHNGA 722
Db 1014 FTGSCQHVYNECDSPCLLGGTCQDGRGLHRCCTCPGGYTGPNCONLVHWCDSPPCKNG 1073
Oy 723 FC-----SAYDECKCTPGMTGLYCTQ-----744
Db 1074 KMGCTHTGY--RCECPSGMTGLYCDVPSVSCVFAAQRGVARLCHGGLCVDAAGTTH 1131
Oy 745 -RCPLEGYGDALY---CO---CQNCAD-DRHSQ--OCTCRGFMGRHCEOK-----788
Db 1132 CRCQAGYTGSGYCDLVDEGSPSPQNGATCTDYLGSGYCKCVAGHGVNCSSEEDCLSH 1191
Oy 789 -----CPSGYGYGCRQICD-----CLNNTC-DRHSQ 815
Db 1192 PCQNGCTCLDLPNTKSCRCRGTGVCENVDCCNPVPDPSRSPCKFNNGTCVDDVG 1251
Oy 816 -TCYSPGMKARGCDQAGVITVGNLN 840
Db 1252 YSCCTCPGPFYGERCE-----GDVN 1270

RESULT 12
AAE18208 standard; Protein; 1473 AA.
XX AAE18208:
XX AC
XX 07-MAY-2002 (first entry)
XX DE
XX Human MOL1b protein.
XX KM Secreted molecule; MOL1b protein; MOLx; cardiomyopathy; atherosclerosis;
XX KM diabetes; chromosomal disorder; albinism; anemia; psoriasis; scarring;
XX KM liver cirrhosis; neurodegeneration; osteoarthritis; organ rejection;
XX KM cerebral thrombosis; hypertension; systemic lupus erythematosus; asthma;
XX KM immune disease; leishmania; immunodeficiency; rheumatoid arthritis; ulcer;
XX KM HIV; human immunodeficiency virus; hepatitis; haemostatic disease; pain;
XX KM hematopoietic; thrombolytic; thrombocytopenia; Alzheimer's disease;
XX KM Parkinson's disease; Huntington's disease; muscular disease; stress;
XX KM ocular disease; growth disorder; depression; epilepsy; contraceptive;
XX KM vulnerability; osteopathy; haemostatic; tranquilizer; antidepressant;
XX KM analgesic; vasodilator; hypotensive; gene therapy; chromosome 1.
XX OS
XX Homo sapiens.
XX FT
XX Key Location/Qualifiers
XX Peptide 1..25
XX FT /label= Signal_peptide
XX Protein 26..1473
XX /label= Mature_MOL1b_protein
XX PN
XX MO200206339-A2.
XX PD
XX 24-JAN-2002.
XX PF
XX 03-JUL-2001: 2001MO-US21249.
XX PR
XX 03-JUL-2000: 2000US-215854P.
XX PR 03-JUL-2000: 2000US-215856P.
XX PR 03-JUL-2000: 2000US-215902P.
XX PR 07-JUL-2000: 2000US-216585P.
XX PR 07-JUL-2000: 2000US-216586P.
XX PR 07-JUL-2000: 2000US-216722P.
XX PR 17-JUL-2000: 2000US-218632P.
XX PR 17-JUL-2000: 2000US-218932P.
XX PR 27-JUL-2000: 2000US-221285P.
XX PR 14-FEB-2001: 2001US-268734P.
XX PA
XX (CURA-) CURAGEN CORP.

```

```

PI Spadena SK, Tchernev V, Liu X, Shenoy S, Spytek K, Zerhusen B;
PI Pattirajan M, Taupier RJ, Rastelli L, Grosse WM, Sekeeres ES;
PI Alsobrook J, Lepley DM, Shen L, Burgess CE, Shinkels RA;
PI Padigaru M;
DR WPI: 2002-155038/20.
DR N-PSDB; AAD28942.
XX
XX Nucleic acids encoding secreted polypeptides, designated MOLx
XX PT polypeptides, useful for treating a MOLx-associated disorder, e.g.
XX PT cardiomyopathy, atherosclerosis, diabetes and metabolic disorders -
XX claim 1; Page 15-16; 223pp; English.
XX
XX The patent discloses nucleic acid sequences encoding novel secreted
XX CC molecule (MOL) polypeptides, designated MOLx polypeptides (i.e. a MOL
XX CC protein where x is an integer from 1 to 8). Sequences of the invention
XX CC are useful for treating or preventing a MOLx-associated disorder in
XX CC humans. They are useful for treating or preventing cardiomyopathy,
XX CC atherosclerosis and disorders related to cell signal processing and
XX CC metabolic pathway modulation. The MOLx antibodies are useful for
XX CC treating or preventing diabetes and disorders related to cell signal
XX CC processing and metabolic pathway modulation. MOLx sequences are useful
XX CC for the treatment or diagnosis of other MOLx-associated disorders, e.g.
XX CC chromosomal disorders, albinism, anaemia, liver cirrhosis, psoriasis,
XX CC scarring, neurodegeneration, osteoarthritis, organ rejection, cerebral
XX CC thrombosis, leishmania, hypertension, systemic lupus erythematosus, immune
XX CC diseases, immunodeficiency, HIV (human immunodeficiency virus), viral,
XX CC bacterial and fungal infections, hepatitis, rheumatoid arthritis, asthma,
XX CC hematopoietic, thrombolytic, haemostatic diseases, thrombocytopenia,
XX CC ulcers, Alzheimer's disease, Parkinson's disease, Huntington's disease,
XX CC ocular disease, muscular diseases, growth disorders, loss of libido,
XX CC stress, depression, pain and epilepsy. They are useful for preventing
XX CC chemotherapy side effects and as contraceptives. Sequences of the
XX CC invention are also useful for gene therapy. The present sequence is
XX CC human Notch-like protein, MOL1b. MOL1b gene is localised on chromosome 1.
XX SQ
XX Sequence 1473 AA:
XX
XX Query Match 15.0%; Score 1014.5; DB 23; Length 1473;
XX Best local similarity 23.5%; Pred. No. 3,3e-45;
XX Matches 339; Conservative 102; Mismatches 361; Indels 643; Gaps 72;
Oy 93 QCCPGEYSEGMKCPHACDKVHR-CIADNTQ-----CEPGMGTCSSACDG 141
Db 91 RCASGF--TGEDCQYSTRSHPCFVSRPCLNGCTHMLSRDYECTCGYGTGKCE-----142
Oy 142 DHNGPHCTSRCCCKNGALCNPIRG--ACHCAAGRGMRCDRECGTYGNDCHQROCON 139
Db 143 -QMTDACLH-PCANGSTCTTVANQFSCKCLTGFGKCE-----TDVNECDIPGHQOR 194
Oy 200 GATCDHYTG--ECRCPPGYTGARCELD-----CPPGKHG 231
Db 195 GGTCLNIPGSGYCCCPGGFTGYCDSLYPCARSPCVNNGSTCQOTGDTFTPECNLPFEG 254
Oy 232 PCCBQ-----RCP-----CONGV 245
Db 255 STCERNIDDCPNHRCQNGVGVGVNNTYNCRCPPMTGQPCFTDVEDCLIPNACONGT 314
Oy 246 CHVTG--ECSRSGMNGTVCGQ-----PPBGRFGNC- 277
Db 315 CANRNGGCGCVGVNMGSGDDCSSEIDDCAFASCTPGSTCIDRYASFCMCPBCKAGLCH 374
Oy 278 -SQEC---OCHNGGTCD--ATGO--CHCSPGYTGERCQ---DECPYGTGVLCATCQC 336
Db 375 LDDACISNPKHAKALCDTNLNGOYICTCPQGTAKGADCTEDVDECA--ANSNPC 427
Oy 327 VNGKCYHSGA--CLCEAGFAGECE-----ARLCPEGLY 360
Db 428 EHAGKCVNTDGAFAHCECLKGYAGPRCEMDINECHSDPCONDATCLDKIGGFTCLCMGFK 487
Oy 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

```

```

Db 488 GVHCELEINECOANPCVNVNQCVDKVNRFQCLCPGFTGVCQIDIDDCSSPTCLNGAKC 547
OY 379 --HPMSGCBACKPGMSGLYCNET-----CSPEYGEAC-QQ 411
Db 548 IDHFNNGECOCATGATGVLCEENIDNCDPDPCHHGQCCODGIDSTYTCICNPGYMGALCSDQ 607
OY 412 I-----CSCONGA-----DCDS---VTG-----K 427
Db 608 IDECYSSPCUNDGRCIDLNVNQCQCPGTGVCNCEINFDCCANPCRHGICMGINRYS 667
OY 428 CTCAPGKGDIC-----STPCPLGYGICSS--RCGCK-----NDAYCS 465
Db 668 CVCSPGFEGGRCNIDIDECASNPCKRGKATCINGVNGFRCLCEPGRHHPSCSYQVNECISN 727
OY 466 P-VDSCT-----CKAGMHGVDCSI-----RCPSGT 490
Db 728 PCINGNCTGGISGYKCLCDDAGWVGINCEVDKNECLSNPCONGTCDNLVNGYRGTCKKG 787
OY 491 WGFPCNLTCQ-----CLNGGAC-----507
Db 788 KGYNCQVNIIDECASNPCLNGTGFDDISGYNCHCVLPYTGKNCQTVLAPSPNPECANAV 847
OY 508 -----NILDGTCTCAPGMRGKCELPCCODGTYGLNCAERCDSSHADGCHPTTGH--CRCL 560
Db 848 CKESPNEFESTYCLCAPGMOGORTYIDIDE-----CISK--PCMHNGLCNHTOGSMECEP 900
OY 561 PGMSGVHCDSCVACGRNGPNCSLPCYCKNGASCSPDGI-----CECAPGFGTTCQR--- 613
Db 901 PFGSGMDEEDIDIDCLANP-----CONGSSC--DGVNTEFCLCPGFGTGDCCQTDNM 951
OY 614 -----ICSPGFYGRCSQTCPOCVHSS-----GPCHTITGL--- 644
Db 952 ECLSEPCKNNGTCDYVNSTYCKQAGFPDGVHCENNNINECESSCFNGTIC--VDGINSF 1009
OY 645 --CDCLPGFTGALC---NEV-----CPSGRFGKNC--AGICT- 674
Db 1010 SCLCPVFTGFSGLHEINECSSHPCLNEGTCVDGLGTYRCSCLPGYTGKNCQTVLNLCSR 1069
OY 675 --CINNNTC--NPIDRSCQCPFGWIGSDCSP-----702
Db 1070 SPCNKNKTCVOKKASQCLCSGNAGAYCDVPNVSCDIASRRGLVEHLCOHSGVCINA 1129
OY 703 -----CPRAHMGPNC---IHTC---NCHNGAFCSAYDG--EECTPGMTGLYCTOR-- 745
Db 1130 GNTHYCCPPLGYTGYSTCEBQDEBECASNPCQHGATCSPFISGRCEYPGYGVANCEYVD 1199
OY 746 -----CPLGFTYG-----KDCALICQCONGADC-DHISGO- 773
Db 1190 EQONQPCONGTCTIDLNVHFKSCSPGTRGILCEENIDCCARGPHCLNGGQCMRIGYS 1249
OY 774 -----CTCRGFMGRCEQKCPSGT 793
Db 1250 CRCLPFRAGERCEEDINECLSNPCSSSEGLDCLQINDYLVCVCSAFTGRICE-----T 1303
OY 794 YGYCROICDCLNNSCTDHTIG-----TCYCSPGMKGARCDQAGVITVINGLNSLRTSTA 848
Db 1304 FVDVCPQM-PELNGTCAVANSMMPDFICRCPFGSGARY--IPEARLRPSVAFPTAM 1359
OY 849 LPADSYQIAGIITILVVLVFLALFLIYRHKOKGESSMPAVTYTPAMRVNADYTI 908
Db 1360 MP---QDDGVAYQITLPAY-----HPFPASVGRYP---TPPSOHSYASSNA 1399
OY 909 SGLTPHNSGANSHTFTNPSTYHTLQ-CATSPHVNNRDRRTVYKSKNNOLFVLKXNVP 967
Db 1400 AERPSSHGALQGHPIYLTSPSPDQWSSSPH-SASDMSDVITSPR-----GGAGG 1452
OY 968 GKRGP 972
Db 1453 GQRGP 1457

```

```

ID AAY06816 standard; protein; 2471 AA.
XX
AC AAY06816;
XX
DT 05-JUL-1999 (first entry)
XX
DE Human Notch2 (humN2) protein sequence.
XX
KW Notch; Nec; Ntn; epidermal growth factor; EGF; repeat domain; cancer;
KW transmembrane domain; cell-fate disorder; proliferative disease;
KW signal transduction; human; Notch2; humN2.
OS Homo sapiens.
XX
FH Key
FT Cleavage-site
FT 25 Location/Qualifiers
FT /note="potential signal cleavage site"
FT 26..143
FT Region
FT /note="EGF repeat region"
FT 1425..1537
FT Region
FT /note="fin-12/notch repeats"
FT 1578..1699
FT Domain
FT /note="transmembrane domain"
FT 1822..2035
FT Region
FT /note="Ankyrin repeats"
FT 2130..2133
FT Region
FT /note="nuclear localisation signal (NLS) sequence"
FT 2149..2152
FT Modified-site
FT /note="putative CKII phosphorylation site"
FT 2156..2159
FT Modified-site
FT /note="putative cdc2 phosphorylation site"
FT 2166..2169
FT Modified-site
FT /note="putative cdc2 phosphorylation site"
FT 2172..2174
FT Region
FT /note="nuclear localisation signal (BNTS) sequence"
FT 2385..2445
FT Region
FT /note="PEST-containing region"
XX
XX W09904746-A2.
XX
XX 04-FEB-1999.
XX
XX 23-JUL-1998; 98MO-US15333.
XX
XX 23-JUL-1997; 97US-0899232.
XX
XX (UYA ) UNIV YALE.
XX
XX Artavanis-Tsakonas S, Qi H, Rand MD;
XX WPI; 1999-153290/13.
XX
XX
XX Measuring and detecting activation of Notch protein in cells from
XX cell surface expression - useful for, e.g. identifying modulators of
XX activation, potentially useful for treating cancer
XX
XX
XX Disclosure; Fig 2A-D; 94pp; English.
XX
XX The invention relates to methods of measuring activation of Notch in a
XX cell. Activated notch can be detected from: (1) expression of: (a) Notch
XX on the cell surface, or (b) one or both of the Notch cleavage products
XX Nec and Ntn; (11) presence of Notch fragments, i.e. one or both of an
XX N-terminal fragment (F1) terminating between the epidermal growth factor
XX -like (EGF) repeat domain and the transmembrane domain or a C-terminal
XX fragment (F2) that has its N-terminus between the specified domains, or
XX fragments of molecular weights about 270, 200, 170, 140, 110, 100, 90 and
XX 85 kDa; or (111) presence of a Notch heterodimer containing a linkage
XX sensitive to reducing agents. The method is used to study or manipulate
XX differentiation processes and to screen/diagnose cell-fate disorders
XX (particularly cancer or other proliferative diseases involving abnormal
XX Notch activation). Modulators of Notch activation are useful for studying
XX or manipulating differentiation and possibly for treating cancer. The
XX F1-F2 heterodimer is the active form of Notch that mediates signal

```



```

FT Region 104..116
FT /note="Immunogenic epitope"
FT Domain 123..138
FT /label=EGF-like_domain
FT Region 136..143
FT /note="Immunogenic epitope"
FT Region 153..159
FT /note="Immunogenic epitope"
FT Region 161..166
FT /note="Immunogenic epitope"
FT Domain 166..181
FT /label=EGF-like_domain
FT Region 200..205
FT /note="Immunogenic epitope"
FT Domain 211..222
FT /label=EGF-like_domain
FT Region 224..230
FT /note="Immunogenic epitope"
XX
XX WO200116156-A1.
XX
XX 08-MAR-2001.
XX
XX 29-AUG-2000; 2000WO-US23663.
XX
XX 30-AUG-1999; 99US-0151348.
XX
XX (HUMA-) HUMAN GENOME SCT INC.
XX
XX Young PE, Ruben SM, NI J;
XX
XX WPI: 2001-183257/18.
XX
XX N-PSDB: AAD02807.
XX
XX Isolated nucleic acids encoding human attractin-like polypeptides,
XX useful for preventing or treating autoimmune diseases e.g. rheumatoid
XX arthritis, and hyperproliferative disorders -
XX
XX Claim 12; Page 219-220; 235pp; English.
XX
XX The present sequence is a HPLIH44 clone human attractin-like
XX protein. Human attractin-like sequences are useful for
XX preventing, treating or ameliorating a medical condition. Human
XX attractin-like sequences are used as a food additive or preservative
XX to increase or decrease storage capabilities and are also useful for
XX chromosome identification and gene therapy. Human attractin-like
XX sequences and their antibodies, agonists and antagonists are useful in
XX the diagnosis, treatment and prevention of cancer such as ovarian cancer,
XX cancer of breast, adrenal gland, bone, bone marrow, breast, liver, lung,
XX gastrointestinal tract, or urogenital system, immune disorders such as
XX Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune
XX thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
XX rheumatoid arthritis and ulcerative colitis, cardiovascular disorders
XX such as myocardial ischaemias, congestive heart failure, pericarditis,
XX wound healing, neurological diseases such as cerebral anoxia,
XX Alzheimer's disease, communicative disorders and epilepsy, and
XX infectious diseases such as viral, bacterial, fungal and parasitic
XX infections.
XX
XX Sequence 241 AA;
XX
XX Query Match 15.0%; Score 1011; DB 22; Length 241;
XX Best Local Similarity 69.1%; Pred. No. 8.7e-46;
XX Matches 152; Conservative 23; Mismatches 45; Indels 0; Gaps 0;
OY 439 CSTPEPLGYINGINSSRCCKNDKAVSPYDGSCTCKAGHGVDCSIRPSGTWGFECNLT 498
DB 5 CAVSGAAGTYGPNSSICSCNNGGTCSPYDGSCTCKEGQGLDCTLPCPSGMLNCNS 64
OY 499 CCGLNGACNTLDGTCTCAPGRGKCELPDGTGGLNCAKRCDCSHADGCHPTTGCR 558
DB 65 CTCANGAACSPIDGSCSTPGLWLDGTCELPDGTGGLNCSHRCDCSHADGCDPYTGHC 124

```

```

OY 559 CLPGMSGVHCDSCVCAEGRWGPCNCSLPCYCKNGKAGSCSPDDGCECAPGFRGTTGCRICSPG 618
DB 125 CLAGWTGICRCDSTCCPRGRGPPNCISVSCSCENSGSCSPEDGSECAPGRGFLCONICPPG 184
OY 619 FYGHRCSQTCPOCVHSSGCPCHITGICDCLPFTGALCNE 658
DB 185 FYGHGAQPCPLCVHSSRCHHISGICELPFTSGALCNG 224
RESULT 15
ABG03826
ID ABG03826 standard; Protein; 587 AA.
XX
XX AC ABG03826;
XX
XX DT 13-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #3817.
XX
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX PA (HXSE-) HXSEQ INC.
XX
XX PI Drmanac RR, Liu C, Tang YT;
XX
XX DR WPI: 2001-639362/73.
XX
XX DR N-PSDB: AAS68013.
XX
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX
XX PS Claim 20; SEQ ID No 34185; 103pp; English.
XX
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 587 AA;
XX
XX Query Match 14.9%; Score 1002.5; DB 22; Length 587;
XX Best Local Similarity 31.0%; Pred. No. 5.7e-45;

```

Matches 214; Conservative 49; Mismatches 178; Indels 249; Gaps 19;

```
OY 31 PIVCSHMEISYTVQESYHPFDQIYYTSTDLNFKTRHRVSRITAYRHGETMYRR 90
Db 5 PTVLCLMAQAYVQ-----RNNHIFWNEKEHG---HG 34
OY 91 KSQCCPGFEYSEGMCPHCADKCVHGRCIAPNTCOCEPGMGNTCS----- 136
Db 35 KSGSC---HNGASC-----SADGACHCTPGWTGLFCTQKPHILASOPLRI 78
OY 137 -----SA 138
Db 79 PCCGLLATVGIQTSREGMQAPGLVVPDSCPTREELCRGSSRPDMIQIDKPKVLQ 138
OY 139 CDGDHMGPHCTSRCKCKNALCNPTTGACHCAAGFRGWRCEDECGTYGNDCHORCOQ 198
Db 139 CPAAFEGKDCGRVCCQNGASCDHISGKCTRTFTGQHCEQRCAPGTFGYCCQQLCECM 198
OY 199 NGATCDHYTGECRCPPGYTGAFCEDE-----LCPPG-----KHGPO----- 233
Db 199 NNSTCDHYTGCTCYCSFGKGRGIRCDGIMLLFLYCAAGPICLASAAEREGPRGSPCL 258
OY 234 ---CEQRCP-----CONGVCHVHTGECSCPSGMMGTVCQ 266
Db 239 LHTCHERXPAPTTPSQDLTDHYLRFESMPIMVLTCLQGA---FPGSPGRPGXTMAPLCGM 314
OY 267 -----POPEGRFGKNSQEQCHNGSGCDAAATGQCHCSPGYTERCODECPVGT 316
Db 315 NVNRPGETHELGDSDHMGPHCSNRCCQNGALCNPTTGACVCAAGFRGWRCEELCAPGTH 374
OY 317 GVLCAETCQCVNGKCYHVSAGCLCEAGFAGERCEARLCPBGLYGIKCDKRCPLHENTH 376
Db 375 GKGCQLPCQCRHGASCDPRAGECLAPGYT-----GY----- 407
OY 377 SCHPMSGECACKPGNSGLYCNETCSPGFYGAQOICSQNGADCDSVTGKCTCAPGFK 436
Db 408 -CHPVTGACTCOPGWSGHHCNESCPSVGYGDCQLPCTCQNGADCHSITGGCTCAPGFMG 466
OY 437 IDGSPCLPGLTYGINCSSRCGCKNDAYCSPVDGSGCTKAGNHGVDCSIKCPSGTWGRG 496
Db 467 EYCAVSCAAGTYGPNCSSTICSNNGGTCTPIDGSCCTKEG--NVP-SLPSPLY----- 518
OY 497 LTCCLNGACNTLDGTCTCAPGWRGEKCELPQDGTGYNLCARCDCHADGCHPTTGH 556
Db 519 -----EHIPVYLPAGE-----SODGTGYNLNCSEHCDCHADGCDPVTGH 558
OY 557 CRCLPGNSGVHCDVCAEG---RWGPNCSL 583
Db 559 CCCLAGWTDIO-----EGFLEKEGPKRTL 582
```

Search completed: May 9, 2003, 11:54:35
Job time : 80.7648 secs

Db 1069 NGKCMQTNNFYKCECKSGMTGYCDVPSVCEVAQAQGVDIYHLCRNSGCVDTGNTH 1128
 QY 443 ---CPLGYGINGSR---CG---CKNDAYSPVDG---SCTCKAGMHGDCS----- 483
 Db 1129 FCRQAGTGSYCEBQVDECPNPNQONATCTDYLGGSCBVAAGHYVNSSEELNCLTS 1188
 QY 484 -----IRCPSTGWFGCNLT---C-----OCLNGAGCNTLDG 512
 Db 1189 HPCQNGGTCTDLIMTYKSCPRGTGVHCEINVDCTPFYDSFTLEPKCFNNKCIDRVG 1248
 QY 513 ---TCTCARGMERKE---LPCQD---GTGLNCAE---RQDC-----SH 546
 Db 1249 GYNCTCPPEGFYGERCEGDVNECLSNPCDSRGTQ---NCIQLVNDYRCBRCRGTGRCEV 1306
 QY 547 ADGC-----HPTTGH---CRCLPGMSGVHCD----- 569
 Db 1307 VDGCKMPCNRGTCVAASNTNERGFICKPFGDGTCEYDSRCSNLRONGGTCTISVL 1366
 QY 570 ---SVCAGRWGPNCSLP 584
 Db 1367 TSSKVCSEGYGTATCQYP 1385

RESULT 4
 US-08-899-232-3
 ; Sequence 3, Application US/08899232
 ; Patent No. 6436650
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Ol, Hulin
 ; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
 ; FILE REFERENCE: 7326-046
 ; CURRENT APPLICATION NUMBER: US/08/899,232
 ; CURRENT FILING DATE: 1997-07-23
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: Patentin Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 2523
 ; TYPE: PRT
 ; ORGANISM: Xenopus sp.
 US-08-899-232-3

Query Match 20.0%; Score 719; DB 4; Length 2523;
 Best Local Similarity 25.8%; Pred. No. 2e-38;
 Matches 222; Conservative 60; Mismatches 222; Indels 355; Gaps 50;

QY 5 LNSLSFLCL-----LICHWIGTASPLLED-----PNVCSHM-----ESYS 41
 Db 603 INELSKRCLNGGCTDRENGYICTCPKGTGVNCTEKIDCASNLCDNGKCIDKIDYE 662
 QY 42 VTVQESYHPFDIYTT-----SCTDIIMWFKCTHRYSTRAYRHGKTYRR 90
 Db 663 CTCRPGTGTGKCLNINIECDSPNCRNGTCKADQJNGFTCV----- 702
 QY 91 KSGCCPGEYSEGENC---VPHC-ADKCVHGR---IAPNTQCEPBGMTNC---SSACD 140
 Db 703 ---CPDGYHD-HMCLSEVNECNSNPCHGACHDGVNGYKDCDCAAGSGSCDINNCE 757
 QY 141 GDHNGPHCTSRCKCNALCPITGA---CHCAAFRGRHCRDREOGTYGDCRQRCQ 198
 Db 758 SN-----PCMNGTCKDMTGAYICTCKAGFSGPNCO---TINIECSSN-PCL 800
 QY 199 NGATC-DHVTG-EFCRCPGYTGAFCEDLCPRGKGGPOEORC---POONGVCHH---V 249
 Db 801 NHGTCIDVAGYKCNCHLPTGALCEAVLAP-----CASSPKNGRCSESEDFE 850
 QY 250 TGECSGSGMGTVCGPCEPGRGKNCSDC---QCHNGTCDATG---QCHSGPYTG 304
 Db 851 TFSCECPGMOGOTC-----EIDMNECVNRPENATQONTNGSYKCNCKPGYTG 900
 QY 305 ERQ---DECPVTVGYVLCATCCQVNGKCYHSGA---CLCEAGFAGEREALD----- 354
 Db 901 RNCMDIDDC-----QPNCHNGSGSCSDGINNFCMCPAGFRPGCEEDINBCAS 950

QY 355 -----CEPHYGIKDRCPCHLE-----NTHSGHPMSG-----DCAC 387
 Db 951 NPKNGANCCTCVNSYNTCTCPGFSGICESENTDCTSSCFNNGTC---IDGINTFTQCQ 1008
 QY 388 KPGWSGLYC---NE-----TCSPEFYDACQOI---CS---CQ 416
 Db 1009 PPGFTGSYCHDINECSKPCPLANGTCQDSYGYKCTCPQGYTGLNQNILVWMDSSPK 1068
 QY 417 NGACCDSTVG---KCTCARGFKIDCSNP----- 442
 Db 1069 NGKCMQTNNFYKCECKSGMTGYCDVPSVCEVAQAQGVDIYHLCRNSGCVDTGNTH 1128
 QY 443 ---CPLGYGINGSR---CG---CKNDAYSPVDG---SCTCKAGMHGDCS----- 483
 Db 1129 FCRQAGTGSYCEBQVDECPNPNQONATCTDYLGGSCBVAAGHYVNSSEELNCLTS 1188
 QY 484 -----IRCPSTGWFGCNLT---C-----OCLNGAGCNTLDG 512
 Db 1189 HPCQNGGTCTDLIMTYKSCPRGTGVHCEINVDCTPFYDSFTLEPKCFNNKCIDRVG 1248
 QY 513 ---TCTCARGMERKE---LPCQD---GTGLNCAE---RQDC-----SH 546
 Db 1249 GYNCTCPPEGFYGERCEGDVNECLSNPCDSRGTQ---NCIQLVNDYRCBRCRGTGRCEV 1306
 QY 547 ADGC-----HPTTGH---CRCLPGMSGVHCD----- 569
 Db 1307 VDGCKMPCNRGTCVAASNTNERGFICKPFGDGTCEYDSRCSNLRONGGTCTISVL 1366
 QY 570 ---SVCAGRWGPNCSLP 584
 Db 1367 TSSKVCSEGYGTATCQYP 1385

RESULT 5
 US-08-400-159-10
 ; Sequence 10, Application US/08400159
 ; Patent No. 5869282
 ; GENERAL INFORMATION:
 ; APPLICANT: Ish-Horowitz, David
 ; APPLICANT: Henrique, Domingos M.P.
 ; APPLICANT: Lewis, Julian H.
 ; APPLICANT: Myatt, Anna M.
 ; APPLICANT: Fleming, Robert J.
 ; APPLICANT: Artavanis-Tsakonas, Spyridon
 ; APPLICANT: Mann, Robert S.
 ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
 ; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036-2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/400,159
 FILING DATE: 07-MAR-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-029
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864

QY 428 CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGK-----NDAYCS 465
 DB 668 CVCSPGFTGQRCNIDIDECASNPCRKATCINGVNGFRCTPBGPHHPSQSYNECLSN 727
 QY 466 P-VGSGCT-----CKAGMHGVDCSI-----RCPSGT 490
 DB 728 PCINGNCTGSLGKCLDADGAVGIVCEVDKNECLSNPCONGTCDNLVNGYRGTCKKGF 787
 QY 491 WGFPCNLTCQ-----CLNGGAC-----507
 DB 788 KGYNQVNIIDECASNPCLNQGTCEDDISGYTCHVLPYTGKNCQTVLAPCSPNCEANAY 847
 QY 508 -----NTLDGTCTCAPGMRGKCEL-----PCQD-----GTV-----GLN 537
 DB 848 CKESPNEFSTYCLCAPGMOGRCITIDIDECISKPCMNHGLCHNTQGSYMCECPGFSQMD 907
 QY 538 CAERCD-----CSHADGCHP--TTGHCRCILPGMSGVHCDSVCAEGRRGPRNC-SLPC 585
 DB 908 CEEDIDCLANPCONGSGCMDGVNTFSCILPFTGDKCQTDNME-----CLSEPC 958

RESULT 10

US-08-083-590A-19
 ; Sequence 19, Application US/08083590A
 ; Patent No. 5786158
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
 ; TITLE OF INVENTION: Nucleic Acids
 ; NUMBER OF SEQUENCES: 21.
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/083,590A
 ; FILING DATE: 25-JUN-1993
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Mistrock, S. Leslie
 ; REGISTRATION NUMBER: 18,872
 ; REFERENCE/DOCKET NUMBER: 7326-015
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212 790-9090
 ; TELEFAX: 212 8698864/9741
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2471 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: peptide
 ; US-08-083-590A-19

Query Match 18.5%; Score 666.5; DB 1; Length 2471;

Best Local Similarity 23.3%; Pred. No. 4.5e-35;

Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

QY 93 QCCPGFYEGEMCPHCAKDCVHGR-CIAPNTCQ-----CEPGMGSTNCSASADG 141
 DB 91 RCASGF--TGEDQCYTSHPCFVSRPCLNGSTCHMLSDTYECTCQVGFTRKEC-----142
 QY 142 DHMGPHCTSRCCCKNGALCNPTIG--ACHCAAGFRGWRCEEDRCBOGTGYGNDCHORCOQN 199

DB 143 -QWTDACLSH-PCANGSTCTTVANQFSCKLGTGTQCE-----TDVNCEDLPHGQH 194
 QY 200 GATCDHTYG--ECRCPPGYTAFCEDL-----CPRGKHG 231
 DB 195 GGTCLNLPSSYQCCPOGFTGOYCDSLVPCAPSPCVNGGTCRQYGTFTPCNCLPREG 254
 QY 232 PQCEQ-----RCP-----CONGV 245
 DB 255 STCERNIDDCPNHRCQNGVCVVDGVNTYNCRCPPQMTGQFCTEDYDECLLPNMQNGT 314
 QY 246 CHAVTG--ECSCPSSMGWTVGQ-----PCBGRGKNC- 277
 DB 315 CANRNGGYGVCVNGMSDDCSENIIDCAFASCTPGSTCIDRVASFSCMCPEGKAGLLCH 374
 QY 278 -SQEC--QCHNGTCDAA--ATGQ--CHCSGYTGERQ--DECPVGTIVLAEYQC 326
 DB 375 LDDACISNPKKALCPTNPLNGYICTCPQGYRGADCTEDYDECBAM-----ANSNPK 427
 QY 327 VNGKCYHVSQA--CLCEAGFAGERCE-----ARLCEGLY 360
 DB 428 EHAKCYVTDGAFHCECLKGTAGRCENDINECHSDPCQNDATCLDKIGFTYCLMFGK 487
 QY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
 DB 488 GVHCELEINECOSNPNVCVNGQCVDKVNRFOCLCPPGFTGYCQIDIDDCSSTPLNGAKC 547
 QY 379 --HMSGCAKCRPMGSLYCNET-----CSRFYEAC-QQ 411
 DB 548 IDHNGYECQCATGTFGLCEBENIDNCDPDCHHGQCODGIDSYTTCICNPGYMAICSDQ 607
 QY 412 I-----CSCQNGA-----DCDS--VTG-----K 427
 DB 608 IDECTSSPCLNDGRCIDLVNGYQCQPGTSGVNCNEINDECAASNPCILHICMGINRYS 667
 QY 428 CTCAPGFKGIDC-----STPCPLGTYGINCSS--RCGK-----NDAYCS 465
 DB 668 CVCSPGFTGQRCNIDIDECASNPCRKATCINGVNGFRCTPBGPHHPSQSYNECLSN 727
 QY 466 P-VGSGCT-----CKAGMHGVDCSI-----RCPSGT 490
 DB 728 PCINGNCTGSLGKCLDADGAVGIVCEVDKNECLSNPCONGTCDNLVNGYRGTCKKGF 787
 QY 491 WGFPCNLTCQ-----CLNGGAC-----507
 DB 788 KGYNQVNIIDECASNPCLNQGTCEDDISGYTCHVLPYTGKNCQTVLAPCSPNCEANAY 847
 QY 508 -----NTLDGTCTCAPGMRGKCEL-----PCQD-----GTV-----GLN 537
 DB 848 CKESPNEFSTYCLCAPGMOGRCITIDIDECISKPCMNHGLCHNTQGSYMCECPGFSQMD 907
 QY 538 CAERCD-----CSHADGCHP--TTGHCRCILPGMSGVHCDSVCAEGRRGPRNC-SLPC 585
 DB 908 CEEDIDCLANPCONGSGCMDGVNTFSCILPFTGDKCQTDNME-----CLSEPC 958

RESULT 11

US-08-532-384-19
 ; Sequence 19, Application US/08532384
 ; Patent No. 6083904
 ; GENERAL INFORMATION:
 ; APPLICANT: Artavanis-Tsakonas, S. et al.
 ; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 ; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
 ; NUMBER OF SEQUENCES: 21
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10036

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-19

Query Match      18.5%; Score 666.5; DB 3; Length 2471;
Best Local Similarity 23.3%; Pred. No. 4.5e-35;
Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

OY 93 OCCPGFYESGEMCVPHCADKCVNHR-CIAPNTCQ-----CEPGNGTNCSSACDG 141
DB 91 RCASGF--TGEDCOYSTHPCFVSRLCLNGTCHMLSRDYTECTOYFTGKEC----- 142
OY 142 DHMGPHCTSRCCCKNGALCNPTG--ACHCAGFRGMRCEDEGCTYGNDCRQCCON 199
DB 143 -QMTDACLSH-PCANGSTCTTVANOFSCKLTGFTGKCE-----TDVNECDIRGHCOH 194
OY 200 GATCDHVTG--ECRCPGYTGAFCEDL-----CPGKHG 231
DB 195 GGTCLNLPSYGCQCPQGTGQYCDLSLYPCASPVCNCGTGRGTGDFTECNCLPGEFG 254
OY 232 POCCEQ-----RCP-----CONGV 245
DB 255 STCERNIDDCPNHRCONGVCVDGVNTYNCRCRPMWTGQFCTEDVDECLLOPNACONGST 314
Y 246 CHHTVG--ECSCPSGMMGVCGQ-----PCPGRGKNC- 277
DB 315 CANRNGIGVCVNGNSGDCSENIDDCAFASCTPGSTCIDRVASFSCMCPGKAGLLCH 374
OY 278 -SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCO---DECPVGTGYVLAETCQC 326
DB 375 LDDACISNCHKGALCDTNPRLNGOYITCPOGYKAGADCTEDVDECAM-----ANSNFC 427
OY 327 VNGKCYHVSQA--CLCEAGFAGERCE-----ARLCEPEGLY 360
DB 428 EHAGKCVNTDGAHFCECLKGAYAGPRCEMDINECHSDPCONDATCLDKTIGFTCLAMPGRK 487
OY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378
DB 488 GVHCELEINECOSNPNVNGOCVDKYNRFQCLCPGFTGTPVCOIDIDDCSSPCLNGAKC 547
OY 379 --HPMSGECACRPGMSGLYCNET-----CSPGYGKNC--QO 411
DB 548 IDHPNGYEQCATGFTGYVLAECENIDNCDPDPCHNGOCODGIDSYTICICPGTNGALCSQ 607
OY 412 I-----CSCQNGA-----DCDS---VTG-----K 427
DB 608 IDECYSSPLNDGRCDLVNGYQCNCQPGTSGVNCIINDDCASNPCINHGICMDGINRS 667

```

```

OY 428 CTCARGFKEIDC-----STPCPLGTYGNCSS--RCGCK-----NDAVCS 465
DB 668 CVCSPEFTGQRKNIDIDECASNPCKRGATPCINGVNGFRICPEGPHHPSYQVNECLSN 727
OY 466 P-VDSCT-----CKAGHGVDCSI-----RCPST 490
DB 728 PCIRHNCITGSLGKCLDAGVNGICEYDKNCKLSNPNQNGTCDNLVNGYRCKKGF 787
OY 491 WFGCNLTQC-----CLNGAC----- 507
DB 788 KGVNCOVINIDECASNPNCLNGTCTFDIDSGTCHVLPYTKNQCQTVLAPCSNPENAMAV 847
OY 508 -----NTLDGTCTCARGRNGEKCEL-----PCOD-----GTY-----GIN 537
DB 848 CKSPNFESEYTCICAPGWGRCCTIDIDECISKPNMGNLCHNTQGSYCECPPEFGSD 907
OY 538 CAERCD-----CSHADGCHP--TTGHCRCLEPGMSGVHCDSCVABGRMGPNC-SLPC 585
DB 908 CEDDIDCLANPCQNGGSCMDGVNTFSCICLPGFTGDKCQTDME-----CLSEPC 958

RESULT 12
US-08-899-232-1
Sequence 1, Application US/08899232
Patent No. 6436650
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
FILE REFERENCE: 7326-046
CURRENT APPLICATION NUMBER: US/08/899,232
CURRENT FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 2471
TYPE: PRT
ORGANISM: Homo sapiens
US-08-899-232-1

Query Match      18.5%; Score 666.5; DB 4; Length 2471;
Best Local Similarity 23.3%; Pred. No. 4.5e-35;
Matches 209; Conservative 64; Mismatches 191; Indels 433; Gaps 46;

OY 93 OCCPGFYESGEMCVPHCADKCVNHR-CIAPNTCQ-----CEPGNGTNCSSACDG 141
DB 91 RCASGF--TGEDCOYSTHPCFVSRLCLNGTCHMLSRDYTECTOYFTGKEC----- 142
OY 142 DHMGPHCTSRCCCKNGALCNPTG--ACHCAGFRGMRCEDEGCTYGNDCRQCCON 199
DB 143 -QMTDACLSH-PCANGSTCTTVANOFSCKLTGFTGKCE-----TDVNECDIRGHCOH 194
OY 200 GATCDHVTG--ECRCPGYTGAFCEDL-----CPGKHG 231
DB 195 GGTCLNLPSYGCQCPQGTGQYCDLSLYPCASPVCNCGTGRGTGDFTECNCLPGEFG 254
OY 232 POCCEQ-----RCP-----CONGV 245
DB 255 STCERNIDDCPNHRCONGVCVDGVNTYNCRCRPMWTGQFCTEDVDECLLOPNACONGST 314
Y 246 CHHTVG--ECSCPSGMMGVCGQ-----PCPGRGKNC- 277
DB 315 CANRNGIGVCVNGNSGDCSENIDDCAFASCTPGSTCIDRVASFSCMCPGKAGLLCH 374
OY 278 -SQEC---QCHNGGTCD--ATGQ--CHCSPGYTGERCO---DECPVGTGYVLAETCQC 326
DB 375 LDDACISNCHKGALCDTNPRLNGOYITCPOGYKAGADCTEDVDECAM-----ANSNFC 427
OY 327 VNGKCYHVSQA--CLCEAGFAGERCE-----ARLCEPEGLY 360
DB 428 EHAGKCVNTDGAHFCECLKGAYAGPRCEMDINECHSDPCONDATCLDKTIGFTCLAMPGRK 487
OY 361 GIKCDKR-----CP-----CHLE-----NTHSC 378

```


SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO: 2
 LENGTH: 2556
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-08-899-232-2

Query Match 18.4%; Score 663; DB 4; Length 2556;
 Best Local Similarity 26.7%; Pred. No. 7.8e-35;
 Matches 204; Conservative 44; Mismatches 192; Indels 324; Gaps 49;

```

QY 94 CCGFYESEMCVPHCADKCVHGR-----IAPPTCCCEPMMGGTNCSSACDGH 143
DB 89 CALGF--SGPLCLPLDNACLTNPCRNGGTCDLLTLEYKRCRCRPGMSKSCQA----- 141
QY 144 WBPHTSRCCCKNGALCNITGA--CHCAAGFRG--WPCEDRCCEG-----TYGNCCHQ- 193
DB 142 --DPCASN-PCANGGCLPEFASITICHCPSPFPGTCMDVNECGKPRLCRHGTCCHNE 198
QY 194 ----RC-----OCNGATC--DHVTGECRCRPGYTGAFC----- 222
DB 199 VGSYRCVCAATHGTPRCCEPVPYPCSPGONGGTCAPTGDVTHCAACLFEGTQNCENI 258
QY 223 DLCPFC--KHGPOC-----EORCP-----CONGVCHHTG- 251
DB 259 DDCPGNNCKNGACVGVNTYNCPCPEPTGOYCTEDVDECOLMPNACONGGTCHHTHG 318
QY 252 --ECSCPSGMWGTVCQ-----PCPEGRFGKNC--SQEC-- 281
DB 319 YNCVAVNGMTGDECSINIDDCASACFHATGCHDRVASFYCECPHRTGLCHLNDACIS 378
QY 282 -OCHNGTCD--ATGO--CHCSPGYTGERCQ--DECPVGTGYVLAETCCQVNGSKCY 333
DB 379 NPCNBSNCDJTPNVNKAICTCPSPGYTGPACSDVDVDECSIGAN-----PCENHAKCI 430
QY 334 HVSGA--CLCEAGFAGEGCEARLCPEGLXGKCDKRC--PCHLENTHSCHPMSGE--CA 386
DB 431 NTLGSECCCLGTYGTPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
QY 387 CRPGWSGLYC-----NE--TCSPGFYGERCQ--QICS--C 415
DB 478 CMPEGYGVHCEVNTDECASSPCLHNGRCIDKINEFQCECPGTGHLCOYDVECASTPC 537
QY 416 QNGADC--DSYTKCTCAGFGKIDCST-----PCPLGTGYGNCSSRCGCKNDAYCS 465
DB 538 KNGAKCLDGPNTTYCTCTGTYGTGHEVDIDECDDPDPCHYGS-----CK-DGVAIT 586
QY 466 PVDSCTCKAGMHGVNCSI--RC--PSGTMGFGCNLCQ-----CLNG----- 504
DB 587 ---FTCLCARPYTGHHCEININECSSQPCRLMG-----TCDDPDNAYLCFCLGTTGPPNC 638
QY 505 -----GAC-NTLDG--TCTCAPGWRGKCELT-----PCODGTY 534
DB 639 ETLNDCCASSPDDSGTCLDKIDYECACEGYTGSNCNSIDECAGNPCHNGTCTCDGIN 698
QY 535 GLNCAERC-----DCSHADGCHPTTGH--CRCLPMSGVNCHD----- 569
DB 699 GFTC--RCPEGYHPTCLSEVNECNSNPVHGACMDSLNGKCDPCPSGNTCDINNNE 756
QY 570 -----SYCAEGRMGPNC-----SLPC 585
DB 757 CESNFCVNGTCKDMTSGIVCTCWEFSGPNCOTINIECASNPC 800

```

RESULT 15
 US-08-083-590A-20
 Sequence 20, Application US/08083590A
 Patent No. 5786158
 GENERAL INFORMATION:
 APPLICANT: Artavanis-Tsakonas, S. et al.
 TITLE OF INVENTION: Therapeutic And Diagnostic Methods
 TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
 TITLE OF INVENTION: Nucleic Acids

```

QY 94 CCGFYESEMCVPHCADKCVHGR-----IAPPTCCCEPMMGGTNCSSACDGH 143
DB 89 CALGF--SGPLCLPLDNACLTNPCRNGGTCDLLTLEYKRCRCRPGMSKSCQA----- 141
QY 144 WBPHTSRCCCKNGALCNITGA--CHCAAGFRGCE--DRCBEG-----TYGNCCHQ- 193
DB 142 --DPCASN-PCANGGCLPEFASITICHCPSPFPGTCMDVNECGKPRLCRHGTCCHNE 198
QY 194 ----RC-----OCNGATC--DHVTGECRCRPGYTGAFC----- 222
DB 199 VGSYRCVCAATHGTPRCCEPVPYPCSPGONGGTCAPTGDVTHCAACLFEGTQNCENI 258
QY 223 DLCPFC--KHGPOC-----EORCP-----CONGVCHHTG- 251
DB 259 DDCPGNNCKNGACVGVNTYNCPCPEPTGOYCTEDVDECOLMPNACONGGTCHHTHG 318
QY 252 --ECSCPSGMWGTVCQ-----PCPEGRFGKNC--SQEC-- 281
DB 319 YNCVAVNGMTGDECSINIDDCASACFHATGCHDRVASFYCECPHRTGLCHLNDACIS 378
QY 282 -OCHNGTCD--ATGO--CHCSPGYTGERCQ--DECPVGTGYVLAETCCQVNGSKCY 333
DB 379 NPCNBSNCDJTPNVNKAICTCPSPGYTGPACSDVDVDECSIGAN-----PCENHAKCI 430
QY 334 HVSGA--CLCEAGFAGEGCEARLCPEGLXGKCDKRC--PCHLENTHSCHPMSGE--CA 386
DB 431 NTLGSECCCLGTYGTPRCEIDV-----NECVSNPC--QNDATCLDQIGEFQCM 477
QY 387 CRPGWSGLYC-----NE--TCSPGFYGERCQ--QICS--C 415
DB 478 CMPEGYGVHCEVNTDECASSPCLHNGRCIDKINEFQCECPGTGHLCOYDVECASTPC 537
QY 416 QNGADC--DSYTKCTCAGFGKIDCST-----PCPLGTGYGNCSSRCGCKNDAYCS 465
DB 538 KNGAKCLDGPNTTYCTCTGTYGTGHEVDIDECDDPDPCHYGS-----CK-DGVAIT 586

```

Query Match 18.4%; Score 662; DB 1; Length 2556;
 Best Local Similarity 26.3%; Pred. No. 9e-35;
 Matches 201; Conservative 46; Mismatches 194; Indels 322; Gaps 46;

NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/083,590A
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Mistrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-015
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 790-9090
 TELEFAX: 212 8698864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2556 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 US-08-083-590A-20

```
OY 466 PYDGSCTCKAGNHGVDCSIR-----CPSGTWFGCNL--- 497
      :|:|:|
DB 587 ---FTCLCRPGYTGHHCEFININECSSQPCRLRGTCQDPDNAYLCFLKGTGPNCIINLD 643
      :|:|:|
OY 498 ---TCCLNGGACNTLDG-TCTCAPGWRGECLE-----PCODGTYGLNCA 539
      :|:|:|
DB 644 DCASSPCDSTCLDKIDGYECACEPgyTGSMSNSNIDECAGNPCHNGGTCEGDGNGFTC- 702
      :|:|:|
OY 540 ERDCSHADGCHPT-----GHCR-----CLPGWSGVHCD----- 569
      :|:|:|
DB 703 -RC---PEGYHDPYCLSEVNECNSNPCVHGACRDSLNGYKCDGPGWSGTINCDIINNNEC 757
      :|:|:|
OY 570 -----SVCABGRWGPNC-----SLPC 585
      :|:|:|
DB 758 ESNPCVNGGTCKDMTSGIYCTCREGFSGPNOCOTNINECASNPC 800
      :|:|:|
```

Search completed: May 9, 2003, 11:59:33
Job time : 24.5435 secs

THIS PAGE BLANK (USPTO)

OY 351 EARLCEGLYGIKDRKCPCHLENTHS 377
DB 249 EARLCEGLYGIKDRKCPCHLENTHS 275

RESULT 3

ABG08033
ID ABG08033 standard; Protein; 878 AA.

AC ABG08033;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #8024.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

BN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: AAS72220.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 38392; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 878 AA;

Query Match 36.4%; Score 1309; DB 22: Length 878;
Best Local Similarity 86.9%; Pred No. 5.5e-55;
Matches 218; Conservative 3; Mismatches 14; Indels 16; Gaps 3;

OY 338 ACLCEAGFAGEARLCEPGLYGIKDRKCPCHLEN--TSHCHPMSEGCACKPGWSGLY 395

DB 5 ALLCQUTVA-----C-----ISAOLICPFANEOQLVACHHPMSGCAKCPGWSGLY 50
OY 396 CNETSGPFGYGGACQOICSCQNGADCDSTYKCTCAPFGKIDGSTPCLGTGYNCSGR 455
DB 51 CNETSGPFGYGGACQOICSCQNGADCDSTYKCTCAPFGKIDGSTPCLGTGYNCSGR 110
OY 456 CGCKNDVAVCSPYDGSCTCAGMHWGVDCTSRCSGTWGFECNLTCCOLNCGACNTLDGCT 515
DB 111 CGCKNDVAVCSPYDGSCTCAGMHWGVDCTSRCSGTWGFECNLTCCOLNCGACNTLDGCT 170
OY 516 CAPGMRGECCELPQCDGTYGLNCAERCDSSHADGCHPTTGHCRCPLPGMSGVHCDVCAEG 575
DB 171 CAPGMRGECCELPQCDGTYGLNCAERCDSSHADGCHPTTGHCRCPLPGMSGVHCDVCAEG 230
OY 576 RWGPNCSLPCY 586
DB 231 RWGPNCSLPCY 241

RESULT 4

ABG22559
ID ABG22559 standard; Protein; 466 AA.

AC ABG22559;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22550.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

BN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: AAS86746.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID NO 52918; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

PI Barnes TM, Fraser CC, Wighton N, Myers P, Busfield SJ, Sharp JD;
XX
XX WPI: 2001-050128/06.
DR N-PSDB; AAF27791.
XX
PT Isolated secreted or transmembrane proteins are used for diagnosis and
PT treatment of neoplastic and haematopoietic disorders e.g. T cell
PT disorders, cancer and tumours -
XX
XX Claim 9; Page 238-240; 294pp; English.
XX
XX The present invention provides the protein and coding sequences for a
CC number of membrane associated and secreted proteins from human, mouse and
CC rat. The proteins are designated INTERCEPT 340, MANGO 003, MANGO 347,
CC TANGO 272, TANGO 295, TANGO 254 and TANGO 378. The proteins are all
CC involved in signal transduction and the sequences can be used in the
CC treatment of cardiovascular, renal, hepatic, neural, neoplastic, skeletal
CC and haematopoietic disorders.
XX
XX Sequence 636 AA;
XX
XX
XX Query Match 33.7%; Score 1215; DB 22; Length 636;
Best Local Similarity 56.7%; Pred. No. 1.2e-50;
Matches 185; Conservative 38; Mismatches 103; Indels 0; Gaps 0;
XX
XX 260 MGTVCQCPCEGRRGKNCSECCNGCTDAATGQCSPGYGECRDECPCVGTGYL 319
DB 1 MGVCISLPCEGRRGKNCSECCNGCTDAATGQCSPGYGECRDECPCVGTGYL 319
XX
XX 320 CAETCQCVNGKCYHVSACLEAGFAGEIRCEARLCPGLYGIKDRKPCHELENTSCH 379
DB 61 CAETCQCVNGKCYHVSACLEAGFAGEIRCEARLCPGLYGIKDRKPCHELENTSCH 379
XX
XX 380 PMSEGCACKGMWSGLYENETCSPEFYGEACQOICSCNGADCDVTKCKCAPEFKIDC 439
DB 121 PMSEGCACKGMWSGLYENETCSPEFYGEACQOICSCNGADCDVTKCKCAPEFKIDC 439
XX
XX 440 STPCPLGTGYNSSRCGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGWFGRGNTLC 499
DB 181 ANLCPPMTYGINSSRCGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGWFGRGNTLC 499
XX
XX 500 QCLNGACNLTDTCTCAPGRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGRCRC 559
DB 241 QCLNGACNLTDTCTCAPGRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGRCRC 559
XX
XX 560 LPGMSGVHCDYVCAEGRGNPCSLPC 585
DB 301 QAGMGTGRCHLPCEGFWGANCNAC 326
XX
XX
XX RESULT 7
ABG08031 ID ABG08031 standard; Protein; 269 AA.
XX
XX ABG08031;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #8022.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX

XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAF72218.
XX
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnosis, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 38390; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
XX
XX Sequence 269 AA;
XX
XX
XX Query Match 32.5%; Score 1169; DB 22; Length 269;
Best Local Similarity 85.6%; Pred. No. 1e-48;
Matches 196; Conservative 3; Mismatches 14; Indels 16; Gaps 3;
XX
XX 338 ACLEAGFAGEIRCEARLCPGLYGIKDRKPCHELENTSCHPMSEGCACKGMWSGLY 395
DB 5 ALICQLTYA-----C-----ISAGLICPFMEQOIVACHPMSEGCACKGMWSGLY 50
XX
XX 396 CNETCSPEFYGEACQOICSCNGADCDVTKCTCAPGFGIDCSTPCPLGTGYNSSR 455
DB 51 CNETCSPEFYGEACQOICSCNGADCDVTKCTCAPGFGIDCSTPCPLGTGYNSSR 455
XX
XX 456 CGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGWFGRGNTLCQCLNGACNLTDTCT 515
DB 111 CGCKNDAYCSPVDSCTCKAGMHGVDCSIRCPSTGWFGRGNTLCQCLNGACNLTDTCT 515
XX
XX 516 CAPGRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGRCRLPMS 564
DB 171 CAPGRGKCELCPCODGTIYGLNCAERDCSHADGCHPTTGRCRLPMS 219
XX
XX
XX RESULT 8
AAG75479 ID AAG75479 standard; Protein; 384 AA.
XX
XX AAG75479;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:6243.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX

PI Young PE, Ruben SM, N1 J;
 XX WPI: 2001-183257/18.
 DR N-PSDB: AAD02810.
 XX
 XX Isolated nucleic acids encoding human attractin-like polypeptides,
 PT useful for preventing or treating autoimmune diseases e.g. rheumatoid
 PT arthritis, and hyperproliferative disorders -
 XX
 PS Claim 12; Page 222-223; 235pp; English.
 XX
 XX The present sequence is a Hf1C0U8 clone human attractin-like
 CC protein. This sequence is encoded by a nucleotide sequence located
 CC on chromosome 1p36.3. Human attractin-like sequences are useful for
 CC preventing, treating or ameliorating a medical condition. Human
 CC attractin-like sequences are used as a food additive or preservative
 CC to increase or decrease storage capabilities and are also useful for
 CC chromosome identification and gene therapy. Human attractin-like
 CC sequences and their antibodies, agonists and antagonists are useful in
 CC the diagnosis, treatment and prevention of cancer such as ovarian cancer,
 CC cancer of breast, adrenal gland, bone, bone marrow, breast, liver, lung,
 CC gastrointestinal tract or urogenital system, immune disorders such as
 CC Addison's disease, allergies, autoimmune hemolytic anaemia, autoimmune
 CC thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis,
 CC rheumatoid arthritis and ulcerative colitis, cardiovascular disorders
 CC such as myocardial ischaemias, congestive heart failure, pericarditis,
 CC wound healing, neurological diseases such as cerebral anoxia,
 CC Alzheimer's disease, communicative disorders and epilepsy, and
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections.
 CC
 XX Sequence 474 AA;
 SQ

Query Match 31.1%; Score 1119.5; DB 22; Length 474;
 Best Local Similarity 44.4%; Pred. No. 3.4e-40;
 Matches 195; Conservative 37; Mismatches 196; Indels 11; Gaps 7;

93 QCCP-GFESEEMCVPHCADKCVHRCIAP--NHCQCEPMGNGTNCSSACDGDHMGPHCT 149
 DB 7 QACPRAGLY--GDNCRHSL--CQNGCTDDPVSGHCACPEBAGNAGLACEHECLPRDYRACR 62
 150 SRQCKNCAALNPITGACHCAGFERWRCEDRCDEGTGTGNDHCRQCCNGATCDHYTGE 209
 DB 63 HSGGCLNGELCDPHYGRCLCPAGWTGDKQSPCLMGFGECACAPRSCCLPAAACHHYTGA 122
 210 CRCPGYGARGCEDLCPGKHGPRCEORCPQ--NGVCHNHTGESCSGMMGYTCGPRC 268
 DB 123 CRCPGFGFGSGCEQACPPGSGFEGECADQCCQCPGENPACHPATGTGSCAAGYHGPSCQORC 182
 269 PEGRGKNCQSEOCCHNGGTDAAMGQCHSPGYTGERCODECPYGTGVCAETCCQCVN 328
 DB 183 PPRHGRPCEDLCCGLNGSCDADATGACRCPTGELGTICNLTCPOGRGPRMTCYHCGGQ 242
 329 GAKCYHVSAGCLCEAGFAGERCEARLCEBGLYGIKCDRCRCHENTHSCPRMSGECACK 388
 DB 243 GAACDPYVGTCLCPRGAGVRCE--RGCPQNRFGVGCETHTSC--RNGELCHANSNCSG 299
 389 PMSGLVQENCTSPGFVEAQQOICSCONGADSDVTGKTCACARFKTIDSTPEPLCTY 448
 DB 300 LGMTRHRELACPPGRGGAACHLSCSNNTCEPATGTCCGPRFYQACEHPCPPFPH 359
 449 GINSRRGCKNDVCSFVDSCTCKAGWHGVDSIRCPSTGFGCNLTCCCLNGAGCN 508
 DB 360 GAGCGGLCMCHGAPCDISGRCLCPAGFHGFCERGCERPSFGEGCHQRDCDGGAPCD 419
 509 TLDTCTCAGMGEKCEL 527
 DB 420 PVTGICLCPGRSGATCNL 438

RESULT 10
 ABG03826
 ID ABG03826 standard; Protein; 587 AA.

XX
 AC ABG03826;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #3817.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-Oct-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB: AAS68013.
 XX
 XX Claim 20; SEQ ID No 34185; 103pp; English.
 PS
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations in
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 XX
 SQ Sequence 587 AA;

Query Match 27.8%; Score 1002.5; DB 22; Length 587;
 Best Local Similarity 31.0%; Pred. No. 1.4e-40;
 Matches 214; Conservative 49; Mismatches 178; Indels 249; Gaps 19;

31 PNVCSHMSVSYVQESFPHFDQIYITSCDILNMFCTHRRSYRATYHGEKTYRR 90
 DB 5 PTVLCMAQAQVGVQ-----RHNIIFWNEKEHG-----HG 34
 91 KSQCCPGFYSEGMCVPHCADKCVHRCIAPNHCCEBGMGTNCS----- 136
 DB 35 KSGSC-----HNGASC-----SABDGACHCTPGWTGLTQTKRPHLLASQPLRI 78
 137 -----SA 138

CC Alzheimer's disease, Parkinson's disease, stroke, immune disorders,
 CC hematopoietic disorders and various dyslipidaemias, metabolic syndrome X
 CC and wasting disorders associated with chronic diseases and cancers,
 CC cardiac disorders, hypertension, hypercalcaemia, cirrhosis, angiodysplasia,
 CC and wound healing, trauma, glomerulonephritis, hyper and hypothyroidism,
 CC multiple sclerosis, lung diseases including asthma, Crohn's disease,
 CC scleroderma, autoimmune diseases, developmental disorders and neural tube
 CC defects. The present sequence is the protein sequence for human MO18b.
 CC MO18b is an acetyl LDL receptor-like protein.
 CC
 CC
 CC

Sequence 865 AA:

Query Match 22.4%; Score 808; DB 23; Length 865;

Best Local Similarity 33.1%; Pred. No. 3.2e-31;

Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

94 CCGPFESEEMC-VPHCA--DKCVHGR-CIANTQCCEPBGKGTNCSACDDDHMPHCT 149
 64 CCAGMRQGDDEGIACVBNSTCSENEVCVRGECRCRGRYGANDCTKCPQFMPGDK 123
 150 SRCCKNGALCNPTGACHCAAGFRGWRCEDEGTGNDCHQRCQNGATCDHVTGE 209
 124 ELCSCHPHQCEDVYGCCTCHA--RRW-----GARCEHACQCHG-TCHEPSGA 169
 210 CCRPGYTGAFCEDLCRPPKHGRQCRPCQNGVCHHTVECSGSGMGTVCQGPCP 269
 170 CRCEPGRMGA-----QCASACYCATSRCDPQTGACICLHAGW----- 206
 270 EGRFGKNSQECQCHNGTCDATGQCCHSPGTGRCODECPVGTGYLCAETCCVNG 329
 207 ---WGRSCNNQCAC-NSSPCEQSGRCQCR-----ER-----TFGARCDRYCCCFRG 249
 330 GRCYHVSACILCEAGFAGERCEARLCPEGLYGIKCDKRCPLHENTHSGPMSEGCACP 389
 250 -----RCHPYDGTACAPR 262
 390 GWSGLYCNETSGRFGYGEACQOIC-SCONGADDSVTGKC-TCARFGKIDSTPCPLGT 447
 263 GYRGYTCRPPCPAGFYGLCRRRCQCKGQPCPTVAEGRLCTCEPGMGTCKDPCATGF 322
 448 YGINSNRG-GCKNDAVCSFVDSGCT-CKAGNHGVDCSIRCSGFWGFCGNLTCCCLNG 505
 323 YEGGSHRCRPPCRDGHACHHTVGTCTRCNAGWIGDRCEKCSNGTYGDECAFYCADCGSG 382
 506 ACNTLDGTCTCAGRWGKEKCELPQDGTGYLCAERCDSSHADGCHPTTGHG 557
 383 HCDPQSGRLCSFVHGPHCNVTCRPLGLHAGDACAQASC-HEDTCDPVTGAC 433

RESULT 14

AAB60393 ID AAB60393 standard; Protein; 866 AA.

AAB60393;

24-APR-2001 (first entry)

Human nurse cell receptor B6TNC#10, SEQ ID NO:21.

Nurse cell receptor; human; B6TNC#10; chromosome 22q11;

DiGeorge's syndrome; drug screening; detection; diagnosis.

Homo sapiens.

JP2000308492-A.

07-NOV-2000.

23-FEB-2000; 2000JP-0045321.

24-FEB-1999; 99JP-0046604.

(SHIO) SHIONOGI & CO LTD.

XX WPI; 2001-161956/17.
 DR N-PSDB; AAF27486.

PT New polypeptide which is a human nurse cell receptor is useful for
 screening assays, especially for DiGeorge's syndrome -

Claim 1; Page 26-28; 48pp; Japanese.

The invention relates to a human nurse cell receptor protein sequence,
 designated B6TNC#10 (AAB60393) and to the cDNA clone encoding it
 (AAF27486). The human B6TNC#10 gene is located on chromosome 22q11,
 in a region associated with DiGeorge's syndrome (a congenital
 disorder which is often associated with heart defects, anomalies of
 the major vessels, oesophageal atresia and abnormalities of facial
 structures). The human B6TNC#10 gene was identified in a human genomic
 sequence (Genbank AC005500) via homology with a murine B6TNC#10 cDNA
 clone (AAF27468) isolated from the murine cell line B6TNC. Appropriate
 PCR primers (AAF27479-AAF27485) were designed to isolate the human
 B6TNC#10 cDNA from the SK-LMS-1 cell line. The invention also relates to
 expression vectors and transformants comprising the human B6TNC#10 cDNA;
 the recombinant production of human nurse cell receptor B6TNC#10; a
 monoclonal antibody recognising B6TNC#10; a hydridoma producing such an
 antibody; a method of screening for a ligand which binds to a human
 nurse cell receptor; and compounds thus identified. The human B6TNC#10
 cDNA can be used in the detection and diagnosis of DiGeorge's syndrome.
 The present sequence represents the human nurse cell receptor
 encoded by cDNA clone B6TNC#10.

Sequence 866 AA:

Query Match 22.4%; Score 808; DB 22; Length 866;

Best Local Similarity 33.1%; Pred. No. 3.2e-31;

Matches 156; Conservative 42; Mismatches 164; Indels 110; Gaps 17;

94 CCGPFESEEMC-VPHCA--DKCVHGR-CIANTQCCEPBGKGTNCSACDDDHMPHCT 149
 64 CCAGMRQGDDEGIACVBNSTCSENEVCVRGECRCRGRYGANDCTKCPQFMPGDK 123
 150 SRCCKNGALCNPTGACHCAAGFRGWRCEDEGTGNDCHQRCQNGATCDHVTGE 209
 124 ELCSCHPHQCEDVYGCCTCHA--RRW-----GARCEHACQCHG-TCHEPSGA 169
 210 CCRPGYTGAFCEDLCRPPKHGRQCRPCQNGVCHHTVECSGSGMGTVCQGPCP 269
 170 CRCEPGRMGA-----QCASACYCATSRCDPQTGACICLHAGW----- 206
 270 EGRFGKNSQECQCHNGTCDATGQCCHSPGTGRCODECPVGTGYLCAETCCVNG 329
 207 ---WGRSCNNQCAC-NSSPCEQSGRCQCR-----ER-----TFGARCDRYCCCFRG 249
 330 GRCYHVSACILCEAGFAGERCEARLCPEGLYGIKCDKRCPLHENTHSGPMSEGCACP 389
 250 -----RCHPYDGTACAPR 262
 390 GWSGLYCNETSGRFGYGEACQOIC-SCONGADDSVTGKC-TCARFGKIDSTPCPLGT 447
 263 GYRGYTCRPPCPAGFYGLCRRRCQCKGQPCPTVAEGRLCTCEPGMGTCKDPCATGF 322
 448 YGINSNRG-GCKNDAVCSFVDSGCT-CKAGNHGVDCSIRCSGFWGFCGNLTCCCLNG 505
 323 YEGGSHRCRPPCRDGHACHHTVGTCTRCNAGWIGDRCEKCSNGTYGDECAFYCADCGSG 382
 506 ACNTLDGTCTCAGRWGKEKCELPQDGTGYLCAERCDSSHADGCHPTTGHG 557
 383 HCDPQSGRLCSFVHGPHCNVTCRPLGLHAGDACAQASC-HEDTCDPVTGAC 433

RESULT 15

AAB60394 ID AAB60394 standard; Protein; 866 AA.

AAB60394;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:54:50 ; Search time 63.1495 Seconds

(without alignments)
853.959 Million cell updates/sec

Title: US-10-092-390-4

Perfect score: 3601
Sequence: 1 MVLINSLCSTICLLCHM1.....HDSVCAEGRWGPNCLSPCV 586

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB pep:*
- 2: /cgn2_6/ptodata/2/pubppaa/PCF_NEW_PUB pep:*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB pep:*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB pep:*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB pep:*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB pep:*
- 7: /cgn2_6/ptodata/2/pubppaa/PCUS_PUBCOMB pep:*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB pep:*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB pep:*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB pep:*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB pep:*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB pep:*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB pep:*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3601	100.0	586	9 US-10-092-390-4	Sequence 4, Appl1
2	3601	100.0	1140	9 US-10-092-390-2	Sequence 2, Appl1
3	1667.5	46.3	1050	9 US-09-796-753-114	Sequence 114, App
4	1215	33.7	636	9 US-09-796-753-100	Sequence 100, App
5	1215	33.7	636	9 US-09-796-753-124	Sequence 124, App
6	1119.5	31.1	474	9 US-10-084-994-11	Sequence 11, Appl
7	843.5	23.4	830	9 US-09-870-759-134	Sequence 134, App
8	843.5	23.4	830	9 US-09-870-759-140	Sequence 140, App
9	843.5	23.4	830	9 US-09-842-758-57	Sequence 57, Appl
10	808	22.4	865	9 US-09-842-758-20	Sequence 20, Appl
11	783.5	21.6	934	9 US-09-842-758-18	Sequence 18, Appl
12	779	21.6	296	9 US-09-866-050A-458	Sequence 458, App
13	779	21.6	299	9 US-09-866-050A-192	Sequence 192, App
14	779	21.6	299	9 US-09-866-050A-332	Sequence 332, App
15	721.5	20.0	310	9 US-10-084-994-12	Sequence 12, Appl
16	664.5	18.5	572	9 US-09-900-449A-7	Sequence 7, Appl1
17	662	18.4	2444	10 US-09-944-849-2	Sequence 5, Appl1
18	656.5	18.2	601	9 US-09-900-449A-5	Sequence 8, Appl1
19	656	18.2	241	9 US-10-084-994-8	

20	645.5	17.9	2167	10 US-09-778-927A-61	Sequence 61, Appl
21	645	17.9	534	9 US-10-125-459-6	Sequence 6, Appl1
22	645	17.9	534	9 US-10-067-761-14	Sequence 14, Appl
23	645	17.9	534	10 US-09-804-156-14	Sequence 14, Appl
24	645	17.9	534	10 US-09-946-633-6	Sequence 6, Appl1
25	644.5	17.9	2201	12 US-10-100-912-2	Sequence 2, Appl1
26	638	17.7	2743	9 US-10-037-182-36	Sequence 36, Appl
27	638	17.7	3695	9 US-10-037-182-2	Sequence 2, Appl1
28	634	17.6	1055	9 US-10-219-248-2	Sequence 2, Appl1
29	634	17.6	1055	9 US-10-219-247-2	Sequence 2, Appl1
30	634	17.6	1055	10 US-09-855-722-2	Sequence 2, Appl1
31	633	17.6	1212	9 US-10-219-248-3	Sequence 3, Appl1
32	633	17.6	1212	9 US-10-219-247-3	Sequence 3, Appl1
33	633	17.6	1212	10 US-09-855-722-3	Sequence 3, Appl1
34	633	17.6	1238	10 US-09-944-849-4	Sequence 4, Appl1
35	630.5	17.5	639	9 US-09-900-449A-4	Sequence 4, Appl1
36	629	17.5	1036	10 US-09-995-593A-6	Sequence 6, Appl1
37	629	17.5	1187	10 US-09-995-593A-7	Sequence 7, Appl1
38	629	17.5	1208	9 US-10-213-329-1	Sequence 1, Appl1
39	629	17.5	1218	10 US-09-995-593A-11	Sequence 11, Appl
40	629	17.5	1238	10 US-09-944-849-3	Sequence 3, Appl1
41	622	17.3	1238	9 US-10-219-248-5	Sequence 5, Appl1
42	622	17.3	1238	9 US-10-219-247-5	Sequence 5, Appl1
43	622	17.3	1238	10 US-09-855-722-5	Sequence 5, Appl1
44	619	17.2	1238	9 US-10-219-248-7	Sequence 7, Appl1
45	619	17.2	1218	9 US-10-219-247-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-10-092-390-4
; Sequence 4, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: No. US20030013865A1 Human EGF-Family Proteins and Polynucle
; FILE REFERENCE: LEX-0311-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIORITY FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 586
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-4

Query Match 100.0%; Score 3601; DB 9; Length 586;
Best Local Similarity 100.0%; Pred. No. 3e-175;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVLINSLCSTICLLCHM1GTASPLNLEDPNVCSSHMSYVYQESTPHRFDDIYTSC 60
DB 1 MVLINSLCSTICLLCHM1GTASPLNLEDPNVCSSHMSYVYQESTPHRFDDIYTSC 60
QY 1 TDLINMFKCTRHRSYRAVHGEKTYRRKSQCCPGFYESGECVPHCADKCHGRCIA 120
DB 1 TDLINMFKCTRHRSYRAVHGEKTYRRKSQCCPGFYESGECVPHCADKCHGRCIA 120
QY 121 PNTQCEPBGWGTNCSSACDGDHMGPHCTSRQCKNKGALCNPTGACCAAGFGWRCED 180
DB 121 PNTQCEPBGWGTNCSSACDGDHMGPHCTSRQCKNKGALCNPTGACCAAGFGWRCED 180
QY 181 RCEGGTGNDCRHCOCQONATGCHVYGECCPGYGTGATCEDCPRGKRGPOEQRCPC 240
DB 181 RCEGGTGNDCRHCOCQONATGCHVYGECCPGYGTGATCEDCPRGKRGPOEQRCPC 240
QY 241 QNGGVCHHYTGECSCPGWMTGCGPCPEGRFRKNCSEQCHNGCTCAATGQCCHCSP 300

```

Db 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP
Qy 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHYSGACLCAGFAGERCEARLCEGLY 360
Db 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHYSGACLCAGFAGERCEARLCEGLY 360
Qy 361 GIKCKRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGFYGEACQOICSCONGAD 420
Db 361 GIKCKRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGFYGEACQOICSCONGAD 420
Qy 421 CDSVGTCKCAGFPGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMVG 480
Db 421 CDSVGTCKCAGFPGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMVG 480
Qy 481 DCSINCPSTGTFGCLNTQCLNGACNTLDGTCTCAPMREKCELPCQDGTGYLNCAL 540
Db 481 DCSINCPSTGTFGCLNTQCLNGACNTLDGTCTCAPMREKCELPCQDGTGYLNCAL 540
Qy 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPCY 586
Db 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPCY 586

```

RESULT 2

```

US-10-092-390-2
; Sequence 2, Application US/10092390
; Publication No. US20030013865A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; TITLE OF INVENTION: No. US20030013865A1el Human EGF-Family Proteins and Polynucleotid
; FILE REFERENCE: Lex-0317-USA
; CURRENT APPLICATION NUMBER: US/10/092,390
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/275,013
; PRIOR FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 2
; LENGTH: 1140
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-092-390-2

```

```

Query Match 100.0%; Score 3601; DB 9; Length 1140;
Best Local Similarity 100.0%; Pred. No. 4,8e-175;
Matches 586; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 WTISLNSCLSTFCLLCHWITGASPLNLEDPNVCSHWESYSTVQDSYHPFDQIYYS 60
Db 1 WTISLNSCLSTFCLLCHWITGASPLNLEDPNVCSHWESYSTVQDSYHPFDQIYYS 60
Qy 61 TDLNMFCTRRVSRVRYRHRGEXTMYRRKSQCCPGAFESGEMCVPHCADCVHGR 120
Db 61 TDLNMFCTRRVSRVRYRHRGEXTMYRRKSQCCPGAFESGEMCVPHCADCVHGR 120
Qy 121 PNTCCPEPMGWTGSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Db 121 PNTCCPEPMGWTGSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Qy 121 PNTCCPEPMGWTGSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Db 121 PNTCCPEPMGWTGSSACDGDHMGPHCTSRQCKNGALCNITGACHCAAGFRNGRCD 180
Qy 181 RCEQGTGNDCHQRCQCGATGCHVTGECRCRPGTGAFCEDLCPGKHGHCQCEGRPC 240
Db 181 RCEQGTGNDCHQRCQCGATGCHVTGECRCRPGTGAFCEDLCPGKHGHCQCEGRPC 240
Qy 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP 300
Db 241 QNGGVCCHHTGECSCPSGMMGTVCQPCPEGRFGKNCSECCCHNGGTDAATGQCHCSP 300
Qy 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHYSGACLCAGFAGERCEARLCEGLY 360
Db 301 GTTGERCODECPVGTGYVLCATCCOCVNGKCYHYSGACLCAGFAGERCEARLCEGLY 360

```

```

Qy 361 GIKCKRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGFYGEACQOICSCONGAD 420
Db 361 GIKCKRCPCHELENTSHCHPMSEGCACKPGMSGLYCNETCSPGFYGEACQOICSCONGAD 420
Qy 421 CDSVGTCKCAGFPGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMVG 480
Db 421 CDSVGTCKCAGFPGIDSTPCPLGTGGINSSRCGCKNDVAVCSVPDSCCTCKAGMVG 480
Qy 481 DCSINCPSTGTFGCLNTQCLNGACNTLDGTCTCAPMREKCELPCQDGTGYLNCAL 540
Db 481 DCSINCPSTGTFGCLNTQCLNGACNTLDGTCTCAPMREKCELPCQDGTGYLNCAL 540
Qy 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPCY 586
Db 541 RDCSHADGCHPTTGHCRCLPGMSGVHCDVCAEGRMGPNCSLPCY 586

```

RESULT 3

```

US-09-796-753-114
; Sequence 114, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; FILE REFERENCE: 7853-227-999
; CURRENT APPLICATION NUMBER: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29

```



```

RESULT 7
US-09-870-759--134
: Sequence 134, Application US/09870759
: Patent No. US2002017551A1
: GENERAL INFORMATION:
: APPLICANT: TERMAN, David S
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASES
: FILE REFERENCE: 870759
: CURRENT APPLICATION NUMBER: US/09/870,759
: CURRENT FILING DATE: 2002-01-14
: PRIOR APPLICATION NUMBER: US 60/208,128
: PRIOR FILING DATE: 2000-05-30
: NUMBER OF SEQ ID NOS: 166
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 134
: LENGTH: 830
: TYPE: prt
: ORGANISM: Homo sapiens
: US-09-870-759-134

```

```

Db 285 EPKMGVQCCQPCLPGRFGESECEQCCCHCKHGEACEPTDTHCQRCDPGLGPCEDPCPT 344
QY 446 GTYGINCSSRCCKNDAVCSVPDSSCTCKAGMHGVDSCIRCPSTGWFGCNLTCCOLNG 505
Db 345 GTFEDDCGSTPCVTQSSCDFTVTGDVCASGYWGPSCNASCAPAGFHGNNCSVPCCEPE-G 403
QY 506 ACNTLDTGTCAPGWR 521
Db 404 LCHPVSGSCQPGSGSR 419

RESULT 8
US-09-870-759-140
: Sequence 140, Application US/09870759
: Patent No. US2002017551A1
: GENERAL INFORMATION:
: APPLICANT: TERMAN, David S
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
: FILE REFERENCE: 870759
: CURRENT APPLICATION NUMBER: US/09/870,759
: CURRENT FILING DATE: 2002-01-14
: PRIOR APPLICATION NUMBER: US 60/208,128
: PRIOR FILING DATE: 2000-05-30
: NUMBER OF SEQ ID NOS: 166
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 140
: LENGTH: 830
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-870-759-140

Query Match 23.4%; Score 843.5; DB 9; Length 830;
Best Local Similarity 36.2%; Pred. No. 7.4e-36;
Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;

QY 93 QCCGGEFESGMC-VPHIC--ADKCVHGR-CLAPTCQCEPBGWGNTGSSACDDHMGPHC 148
Db 40 QCCAGMWKQKDECTIPICEBPDAQCXDEVCVKPELCKRKFGEAHCSRCPQYWGPD 99
QY 149 TSPROCCKNKGALNPITGACHCAAGFRGMRCEDRCQGTGYNDCORCOCONGATCDHYTG 208
Db 100 RESPCRPHHGCCEBATATACCCQADRMGARCEFFCACGPHR-----CDPATG 146
QY 209 ECRCPPEYTGAFCEDLCPKGNHGPCEBRCRPPCONGVCHVHTGECSCSPSGMGTVCGRPC 268
Db 147 VCHGEPMWSTYCRRCPCQCNMTAARCEQ-----ATGACVCKRPGM----- 185
QY 269 PEGRFGKNCSECCNHGGTCDAAATGQCCHSCSPGTGERCODECPVGTGYLAETGCYV 328
Db 186 ---WGRCSFRNCMH-GSPCEDSGSRACRPGMWGPECCQO-----CECYR 227
QY 329 GSKCYHVSAGACLAGAGFAGRCEARLCPGLYGTCKDKRRCPLLENTHSCHPMMSGRC-AC 387
Db 228 -GRCSAASGECTCPGFRGANCELP-CPAGSHGQCAHSCG-RCKNHPEPSPDTGSCESC 284
QY 388 KPGMSGLYCNETCSPGFYGEACQOIC-SCONGADCDSTVTKGK-TCAAPFGKIDSTPCPL 445
Db 285 EPGMNGVQCCQPCLPGRFGESECEQCCCHCKHGEACEPTDTHCQRCDPGLGPCEDPCPT 344
QY 446 GTYGINCSSRCCKNDAVCSVPDSSCTCKAGMHGVDSCIRCPSTGWFGCNLTCCOLNG 505
Db 345 GTFEDDCGSTPCVTQSSCDFTVTGDVCASGYWGPSCNASCAPAGFHGNNCSVPCCEPE-G 403
QY 506 ACNTLDTGTCAPGWR 521
Db 404 LCHPVSGSCQPGSGSR 419

RESULT 9
US-09-842-758-57
: Sequence 57, Application US/09842758
: Publication No. US20030083244A1
: GENERAL INFORMATION:

```



```
APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R
APPLICANT: Gerlach, Valerie
APPLICANT: Shinkets, Richard A
APPLICANT: Malyankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigaru, Muralidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esna A
APPLICANT: Smithson, Glenda
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,508
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/220,591
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/232,678
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/263,217
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 60/265,160
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 57
LENGTH: 830
TYPE: PRT
ORGANISM: Homo sapiens
US-09-842-758-57

Query Match      23.4%  Score 843.5; DB 9; Length 830;
Best Local Similarity 36.2%  Pred. No. 7.4e-36;
Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;
```

```
QY 93 QCCPFYSESGMC-VPHC-ADKCVHGR-CIAPNTCCCEPFGGTTNCSASACDGDHMGPHC 148
DB 40 OCCAGMRKDDOCTCTIPICBGPACCKDEVCVAPGLCRCKPFGFAHCSRCFGOTWGPCC 99
QY 149 TSCCKCKNGALCNPTTGACHCAAGFRGWRCEDECRCEQGTGYNDCDHORCCQNGATCDHYTG 208
DB 100 RESCCHPHGCEPATGACCCADRMWARGCEPFCACGPHGR-----CDPATG 146
```

```
QY 209 ECRCPGTYGAFCEBCLCPGKGHGPOCEORCPONGCVCHHTGECSCSPSGMWGTYCGGPC 268
DB 147 VCHCEPGRWMSYCRBRPCCQNTAAARCE-----ATGACVCKRPGW----- 185
QY 269 PEGRFGKNSQOCCHNGTCAATGCCCHCSFGYTGERRQDDCPRTGYVLCATFCQCN 328
DB 186 ---WGRRCSEFRCNH-GSPCEQDSGRACRGMWGPCEQDQ-----CECVR 227
QY 329 GGRKYHVSAGCAGCEAGFAGERCEARLCPGLTGICDKRCPLHTSHCHPMSECC-AC 387
DB 228 -GRCSAASEGCTCPGFRARCELP-CPAGSHGVCAHSCG-RCKHNPCSDPTGSCSC 284
QY 388 KPGMSGLYCNEFCSPGFFGEACQITC-SCQNDADDSTYTKC-TCAPFGKIDCSTPCL 445
DB 285 EPGWNGTCCQCPCLPPTGTFESCEQOCPICRHGEACEPDTGHCQRCDPGLGPRCEPCPT 344
QY 446 GTYGINCSRGCKNDVAVSPYDSCTCKAGHWGDCSIRCPSGTWGFCNLTCCLNG 505
DB 345 GTFEGDCSTCTPTCYQGSCTDVTYGDVCSAGTYWPSGSCNCPAGFRHNNCSVPECEPE-G 403
QY 506 ACNTLDGTCTCAPGWR 521
DB 404 LCHPVSGSQPGSGSR 419

RESULT 10
US-09-842-758-20
Sequence 20, Application US/09842758
Publication No. US20030083244A1
GENERAL INFORMATION:
APPLICANT: Vernet, Corine A. M.
APPLICANT: Fernandes, Elma R
APPLICANT: Gerlach, Valerie
APPLICANT: Shinkets, Richard A
APPLICANT: Malyankar, Uriel M
APPLICANT: Boldog, Ferenc L
APPLICANT: Zernusen, Bryan D
APPLICANT: Spytek, Kimberly A
APPLICANT: Majumder, Kumud
APPLICANT: Tchernev, Velizar T
APPLICANT: Padigaru, Muralidhara
APPLICANT: Patturajan, Meera
APPLICANT: Burgess, Catherine E
APPLICANT: Gangolli, Esna A
APPLICANT: Smithson, Glenda
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758
CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
```

```

1  PRIOR APPLICATION NUMBER: 60/201,508
2  PRIOR FILING DATE: 2000-05-03
3  PRIOR APPLICATION NUMBER: 60/220,551
4  PRIOR FILING DATE: 2000-07-25
5  PRIOR APPLICATION NUMBER: 60/233,678
6  PRIOR FILING DATE: 2000-09-15
7  PRIOR APPLICATION NUMBER: 60/263,217
8  PRIOR FILING DATE: 2001-01-32
9  PRIOR APPLICATION NUMBER: 60/265,160
10 PRIOR FILING DATE: 2001-01-30
11 NUMBER OF SEQ ID NOS: 113
12 SOFTWARE: PatentIn Ver. 2.1
13 SEQ ID NO 20
14 LENGTH: 865
15 TYPE: PRT
16 ORGANISM: Homo sapiens
17 JS-09-842-758-20

```

Query Match	22.48:	Score 808;	DB 9;	length 865;
Best Local Similarity	33.18:	Pred. No. 4.7e-34;		
Matches 156;	Conservative 42;	Mismatches 164;	Indels 110;	Gaps 17;

QY	CCPFGYESGEMC -VPHCA -DKCYHGR -CIAAPNCOEPEPMGGNCSACDGDHMGHCT	14.9
Db	64 CCAGRRQGGDCGCAVCEGNSSTCSENEVCYARPECRNRHGFPGANCTKCPROFGEPODK	123.3
QY	150 SRCOCKNGALCNPTTGACHCAAGFERGWRCEDEBOGTGYNDCHARRCCOQNGATCDHYTGE	209
Db	124 ELCSCHPHGQCEEDVTGQOTCHA -RRW-----GARCEHAQQCQHG -TCHPRSGA	169
QY	210 CRCPFGYTGACBEDLCPPGKKHGPCCEORCPQNGGVCHNYTGBESCSGSMNGYVQGP	268
Db	170 CRCEGWMGA-----QCASACTSATSIRDPTGALCLHAGW-----	206
QY	270 EGRFGKNSOECOCNHNGTCAATGAGCHCSBSPHYGGERCODEPCVGTGYVLAETQCQVNG	329
Db	207 ---WGRSNNQAC -NSSPCGQDSGRQGR---ER-----TIGACDRRCQCFRG	24.9
QY	330 GKCYHVSAGCLCEAGFAGERCEARLCPBLGYIKDRCRCPCHLENTHSCHPMGECAKCP	369
Db	250 -----RCHPVGTGCAEP	262
QY	390 GMSGLYCNETCSPEFYEGACQOLC -SGONADCDQSVTGYKC -TCAPEGKIGDSTPCLGT	44.7
Db	263 GTRGKYTCEPCPAGFYTGAGCRRRCGQCGQDPCTVAABGRCLTCEPBGNGTKCDQPCATYGF	322
QY	448 YGINCSSRC -GCKNDAYCSPYDVSCT -CKAGWHGVDSIRSPPSTWGFGLNTLCOCLNGG	50.5
Db	323 YGEGCSHRCPCPCRDGHACNHVTGCTCRKMGWIGIDRCETKCSNGETYGEDCAFVQADCGSG	382
QY	506 ACNTLIDGCTCTAPGWRBKEKCLPQDQGTGYLNCABERDCSHADGCHPTTGHG	55.7
Db	383 HCDPSGKCLTSPVHGPHCVNTPCPGLHGDADQAOQSC -HEIDCDPVTGAC	43.3

US-RESULT 11
 US-09-842-758-18
 : Sequence 18, Application US/09842758
 : Publication No. US20030083244A1
 GENERAL INFORMATION:
 : APPLICANT: Vernet, Corinne A. M.
 : APPLICANT: Fernandes, Elma R.
 : APPLICANT: Gerlach, Valerie
 : APPLICANT: Shinkens, Richard A
 : APPLICANT: Malvanekar, Uriel M
 : APPLICANT: Boldog, Ferenc L
 : APPLICANT: Zernusen, Bryan D
 : APPLICANT: Spytek, Kimberly A
 : APPLICANT: Majumder, Kunind
 : APPLICANT: Tchenev, Velizar T
 : APPLICANT: Padigara, Muralidhara
 : APPLICANT: Patirajan, Meera
 : APPLICANT: Burgess, Catherine E

```

APPLICANT: Gangolli, Esha A
APPLICANT: Smithson, Glenda
APPLICANT: Rastelli, Luca
APPLICANT: MacDougall, John R
APPLICANT: Taupier, Raymond J
APPLICANT: Grosse, William M
APPLICANT: Edward, Szekeres S
APPLICANT: Alsobrook II, John P
FILE OF INVENTION: No. US20030083244A1 Proteins and Nucleic Acids Encoding Sam
FILE REFERENCE: 15966-783
CURRENT APPLICATION NUMBER: US/09/842,758
CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/200,158
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: 60/200,613
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/200,780
PRIOR FILING DATE: 2000-04-28
PRIOR APPLICATION NUMBER: 60/201,006
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,007
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,236
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,238
PRIOR FILING DATE: 2000-05-01
PRIOR APPLICATION NUMBER: 60/201,186
PRIOR FILING DATE: 2000-05-02
PRIOR APPLICATION NUMBER: 60/201,474
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/201,508
PRIOR FILING DATE: 2000-05-03
PRIOR APPLICATION NUMBER: 60/220,591
PRIOR FILING DATE: 2000-07-25
PRIOR APPLICATION NUMBER: 60/232,678
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 60/263,217
PRIOR FILING DATE: 2001-01-22
PRIOR APPLICATION NUMBER: 60/265,160
PRIOR FILING DATE: 2001-01-30
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 934
TYPE: PRT
ORGANISM: Homo sapiens
US-09-842-758-18

```

Query Match	21.8%	Score 783.5	DB 9	Length 934
Best Local Similarity	31.7%	Pred. No. 8.6e-33		
Matches 160; Conservative	48	Mismatches 190	Indels 107	Gaps 19

Qy	94	CCPPEFEGEMC-VPHCA--DKCHNR-CLAPNTCOEPBGMGNSSACD-----	14.1
Db	64	CCAGRRQGDDECGIAYCEGNSSTCSENECYARGECRKHGFFGANCJTSERGVLYVG	12.3
Qy	142	-DHW-----GRHCTSRCCCKNGALCNPTTGAOCHCAAGRGW	17.6
Db	124	AESWRDGAQSKVGRGRTIRLGGSPDEVAAGVADRAPFRLAGTYS--TGAFH---PLRSS	17.8
Qy	177	RCEHRCBEGTGYGNDCHORCOONCATDHWYGEBCRCRPTGTGACBDELCPRKHGPPCEQ	23.6
Db	179	PAE--CRPFQFPPCKELCSCNHPGOCEDYTGQCTCAARRGA-----RCEH	22.3
Qy	237	RCPCQNGVCNHHYTGECSPBGMGTGCGORCPBGRFGKNCQOBCQCHNGTCDATGQC	23.6
Db	224	ACOCOH-GTCHPRSGACBCEGSM-----WGACASACYSATSRCDDPQTGAC	26.9
Qy	297	HCSGTGYGERQDDCPRVGTGYLVLAETGQCQCVNNGKCYHVNAGACILEAGFAGBCEARLCP	35.6
Db	270	LCHGMMWRGSRNNOC-----ACNSSPCDDQSGRCCR-----	30.1
Qy	357	EGLYGKICKDKRCRPHLENTWTHSCHPMSBECAPKPMGSLYCNETICSPGTYEACQDIT-SC	41.5

Db 302 ERTFARADRYCOF---RGRCHPYDGTACBPRTGKTYCRKCPACAGTGLGCKRRCCGC 358
OY 416 QNGADCDSDYTKC-TCAPDFKIDCSTPCPLDTYGINCSSRC-GCKNDVAVSPVDGSGT 472
Db 359 KQODCTVAEGHCLTCEPDMNKTCKDQPCATGTYEGEGCSHRCPORDGHACHVTKCTR 418
OY 473 CKAHGHVDCSIRCPSTGTMGFCNLTQCLNAGACNTLDGTCTCAPGRGECGLPCODG 532
Db 419 CNAIGTIDRCETKCSNGTYGEDCAFVADCGSGHCDPQSGRCLCSPGVHGRHCNTYCPBG 478
OY 533 TYGLNCAERCDOSHADGCHPTTGHG 557
Db 479 LHGADCAQACSC-HEDTCDPYTGAC 502

RESULT 12

US-09-866-050A-458
; Sequence 458, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Compositions and Methods for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 458
; LENGTH: 296
; TYPE: PRG
; ORGANISM: Mouse
US-09-866-050A-458

Query Match 21.6%; Score 779; DB 9; Length 296;
Best Local Similarity 39.9%; Pred. No. 6.6e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

OY 165 GACHCAAGFRMGRCEDRCOEGTYGNDCHORCOCONGATCDHVTGECRCPPGTGAFEDL 224
Db 1 GACYCPRAGFLGADCSLACPOGRFPSCAHVTCGGAACDPVSTGTCPCPKTGHCERG 60
OY 225 CPPEKHGPOCEORPCQNGVCHVTGECSCPSGMGTVCQPCPEGRFGKNCQSECOCH 284
Db 61 CPDRFPGKCEHCKACRNGGLCHATNGSCSCLPMGMGRPHCHNACPARHYGAACLLCSCQ 120
OY 285 NGCTCDATGQCHSPYTGRCDECPVGTGYVLAETCCVNGKCYHVSAGCLCEAG 344
Db 121 NNGSCPTSGACLCGPFGYQACEDTFCPAGHSGSCQRVCECQGAACDPVSGRCLCPAG 180
OY 345 FAGERCEARLCPBEGLYGIKDKRCHLNTSHCHPMHSGEACRPGMSGLYCNCTGSPGF 404
Db 181 FRGO-----FCERGCPRGF 194
OY 405 YGACQOICSONGADSDYTKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAYC 464
Db 195 FGDGCLQOQNCPTGVPDPIISGLCLCPRGRAGTTCDLDCRGRGRGPCALRCDCGGGADC 254
OY 465 SPVDGSGCTCAGMHGVDC 482
Db 255 DPISGQCHCVDSYTGPTC 272

RESULT 13

US-09-866-050A-192
; Sequence 192, Application US/09866050A
; Publication No. US20030040471A1

; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Compositions and Methods for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; LENGTH: 299
; TYPE: PRG
; ORGANISM: Mouse
US-09-866-050A-192

Query Match 21.6%; Score 779; DB 9; Length 299;
Best Local Similarity 39.9%; Pred. No. 6.7e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

OY 165 GACHCAAGFRMGRCEDRCOEGTYGNDCHORCOCONGATCDHVTGECRCPPGTGAFEDL 224
Db 4 GACYCPRAGFLGADCSLACPOGRFPSCAHVTCGGAACDPVSTGTCPCPKTGHCERG 63
OY 225 CPPEKHGPOCEORPCQNGVCHVTGECSCPSGMGTVCQPCPEGRFGKNCQSECOCH 284
Db 64 CPDRFPGKCEHCKACRNGGLCHATNGSCSCLPMGMGRPHCHNACPARHYGAACLLCSCQ 123
OY 285 NGCTCDATGQCHSPYTGRCDECPVGTGYVLAETCCVNGKCYHVSAGCLCEAG 344
Db 124 NNGSCPTSGACLCGPFGYQACEDTFCPAGHSGSCQRVCECQGAACDPVSGRCLCPAG 183
OY 345 FAGERCEARLCPBEGLYGIKDKRCHLNTSHCHPMHSGEACRPGMSGLYCNCTGSPGF 404
Db 184 FRGO-----FCERGCPRGF 197
OY 405 YGACQOICSONGADSDYTKCTCAPGFKGIDCSTPCPLGTYGINCSSRCGCKNDAYC 464
Db 198 FGDGCLQOQNCPTGVPDPIISGLCLCPRGRAGTTCDLDCRGRGRGPCALRCDCGGGADC 257
OY 465 SPVDGSGCTCAGMHGVDC 482
Db 258 DPISGQCHCVDSYTGPTC 275

RESULT 14

US-09-866-050A-332
; Sequence 332, Application US/09866050A
; Publication No. US20030040471A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Ornust, Rene
; APPLICANT: Murison, James G.
; TITLE OF INVENTION: Compositions and Methods for Isolated From Skin Cells
; FILE REFERENCE: 11000.1011c4U
; CURRENT APPLICATION NUMBER: US/09/866,050A
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 725
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 332
; LENGTH: 299
; TYPE: PRG
; ORGANISM: Mouse
US-09-866-050A-332

Query Match 21.6%; Score 779; DB 9; Length 299;
Best Local Similarity 39.9%; Pred. No. 6,7e-33;
Matches 127; Conservative 27; Mismatches 118; Indels 46; Gaps 1;

QY 405 YGEACQOICSCONGADCDSTYTKCTCAPGFKGIDCSTPCPL 445
DB 238 FGPSCTLHCDGCGGADCDPVSGQCHVDYMGPTCRREGPL 278

Search completed: May 9, 2003, 12:02:22
Job time : 68.1495 secs

QY 165 GACHCAAGFRKRCEDRCEDGTGNDCHQRCQCONGATCDHYTGECRCPREYTGAFCEDL 224
DB 4 GACTCPAGFLCADCSLACPGREFRPGSCAHVCTCGGAACDPVSGTCLCPGKGTGHCERG 63
QY 225 CPPKRGPOCEQRCPONGGVCHVHTGECSPSGMMTVGQPCPEGRFGNCSQEQCH 284
DB 64 CPDPRFGKCHRCACRNNGSLCHATNGSCSCLGMMGPCEHACPAGRYGAACLLCECSCQ 123
QY 285 NGGTCDATGQCHCSPGYTGECRDECPVGTGYVLAETCQCVNGKCYHVSAGCLCEAG 344
DB 124 NNSCEPTSGACLCGPFYTGACEDTTPAGFHSGCQRCVCECQOGAPCDPVSGRLCPAG 183
QY 345 FAGERCEARLCPBGLYIKDKRCRCHLENTHSCHPMSGECACRPGWSGLYCNETCSPGF 404
DB 184 FRGQ-----FCERCKRGF 197
QY 405 YGEACQOICSCONGADCDSTYTKCTCAPGFKGIDCSTPCPLGTGICNSSRCCGKNDAYC 464
DB 198 FGDCGLDQCCNCPTEVPDPISTGLCLCPPRAGTTCDLDCRRGRFGPCALRCDCGGADC 257
QY 465 SPVDSCTCKRAGMHGVC 482.
DB 258 DPISGQCHVDSTYTGPTC 275

RESULT 15
US-10-084-994-12
; Sequence 12, Application US/10084994
; Publication No. US20030023070A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: Attractin-like polynucleotides, polypeptides, and antibodies
; FILE REFERENCE: PTO11P1C1
; CURRENT APPLICATION NUMBER: US/10/084,994
; PRIOR FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/790,621
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: PCT/US00/23663
; PRIOR FILING DATE: 2000-08-29
; PRIOR APPLICATION NUMBER: US 60/151,348
; PRIOR FILING DATE: 1999-08-30
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 12
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-084-994-12

Query Match 20.0%; Score 721.5; DB 9; Length 310;
Best Local Similarity 44.5%; Pred. No. 5.5e-30;
Matches 125; Conservative 24; Mismatches 129; Indels 3; Gaps 2;

QY 165 GACHCAAGFRKRCEDRCEDGTGNDCHQRCQCONGATCDHYTGECRCPREYTGAFCEDL 224
DB 1 GPAAPLGSSTRTATSPVPOGRFPGNCTHYGCGGACADPVTTCTCLCPGRACVHCERG 60
QY 225 CPPKRGPOCEQRCPONGGVCHVHTGECSPSGMMTVGQPCPEGRFGNCSQEQCH 284
DB 61 CPQNRFGVGEHNTSCNRNGSLCHANSNGSCGGLWTRHCEIACPPGRITGAACHLECSCH 120
QY 285 NGGTCDATGQCHCSPGYTGECRDECPVGTGYVLAETCQCVNGKCYHVSAGCLCEAG 344
DB 121 NNSCEPTSGACLCGPFYTGACEDTTPAGFHSGCQRCVCECQOGAPCDPVSGRLCPAG 180
QY 345 FAGERCEARLCPBGLYIKDKRCRCHLENTHSCHPMSGECACRPGWSGLYCNETCSPGF 404
DB 181 FHGHCE-RGCEPSFEGCHQRCDC--DGAPCDPVTTGLCLCPPRSGATCNLDCCRGO 237

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 11:47:55 ; Search time 16.9757 seconds

(without alignments)
3318.560 Million cell updates/sec

US-10-092-390-4

Sequence: 1 MYSLNSCSFICLLICHLI.....HDSVCAGRWGNCSLPcy 586

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database 8
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1372.5	38.1	1574	2 T13954	MEGF6 protein - ra
2	1322	36.7	1620	2 T22783	hypothetical prote
3	1284.5	35.7	1111	2 T26872	hypothetical prote
4	719	20.0	2524	2 A35844	notch protein - Af
5	717	19.9	2437	2 S42612	transmembrane prot
6	697	19.4	2321	2 S78549	notch3 protein - h
7	693	19.2	2531	2 A46019	Notch-1 protein -
8	687	19.1	1203	2 A49175	Notch B protein -
9	685.5	19.0	2318	2 S43506	notch 3 protein -
10	685.5	19.0	2471	2 A49128	cell-fate determin
11	682	18.9	3566	1 A40701	tenascin-X precurs
12	681.5	18.9	2531	2 T31070	notch homolog - se
13	677	18.8	2703	1 A24420	notch protein - fr
14	676.5	18.8	4206	2 T09070	probable tenascin
15	675	18.7	2531	2 S18188	notch protein homo
16	672.5	18.7	4135	2 T42629	tenascin-X - Bov
17	664.5	18.5	1064	2 A40136	fibropellin Ia - s
18	659.5	18.3	2555	2 A40043	notch protein homo
19	658	18.3	1964	2 T09059	notch4 - mouse
20	648.5	18.0	2352	2 T30201	Notch homolog prot
21	644.5	17.9	2201	2 A32160	tenascin-C - human
22	640.5	17.8	2019	1 J01322	tenascin precursor
23	636	17.7	1810	1 A32230	tenascin precursor
24	631.5	17.5	2139	2 A35672	crumbs protein - f
25	616	17.1	1220	2 A56136	tenascin precursor
26	611	17.0	1746	1 S19694	hypothetical prote
27	593.5	16.5	3672	2 T23433	probable laminin a
28	593.5	16.5	3704	2 T37316	gene serrate prote
29	587	16.3	1408	2 S16148	

30	586	16.3	1722	2 E89753	protein FL1C7.4 (1
31	586	16.3	3635	2 T10053	laminin alpha 5 ch
32	580.5	16.1	861	2 A48825	Notch homolog Mote
33	577.5	16.0	1801	1 MMR1S	laminin beta-2 cha
34	576.5	16.0	3712	2 S18253	laminin alpha-1 ch
35	565.5	15.7	2823	2 T23064	hypothetical prote
36	565.5	15.7	2823	2 F87908	protein T22A3.8 (1
37	565.5	15.7	3102	2 T43291	laminin alpha chal
38	562.5	15.6	647	2 A43802	tenascin - eastern
39	561	15.6	1798	2 S53869	laminin beta-2 cha
40	561	15.6	3106	1 S53868	laminin alpha-2 ch
41	560.5	15.6	1429	2 S06434	homeotic protein 1
42	556	15.4	833	2 S19087	gene Delta protein
43	552	15.3	832	2 A31246	neurogenic protein
44	552	15.3	880	2 S00670	neurogenic repeat
45	550	15.3	1797	2 A53677	laminin beta-2 cha

ALIGNMENTS

RESULT 1					
T13954					
MEGF6 protein - rat					
C:Species: Rattus norvegicus (Norway rat)					
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000					
C:Accession: T13954					
R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.					
Genomics 51, 27-34, 1998					
A:Title: Identification of high-molecular-weight proteins with multiple EGF-like motif					
A:Reference number: Z14126; M0ID:98360089; PMID:9693030					
A:Accession: T13954					
A>Status: preliminary; translated from GB/EMBL/DBD					
A:Molecule type: mRNA					
A:Residues: 1-1574 <MAX>					
A:Cross-references: EMBL:AB011532; NID:93449293; PIDN:BA32462.1; PID:93449294					
A:Experimental source: strain Sprague-Dawley; brain					
C:Genetics:					
A:Gene: MEGF6					
Query Match					
Best Local Similarity 41.3%; Pred. No. 7.4e-69;					
Matches 229; Conservative 56; Mismatches 198; Indels 71; Gaps 9;					
QY	94	CCGFTYSGEMCYPHCADCVHG-----RCIAPN-----TCGEPGSGTNC	136		
DB	808	CLPEFVGS-----RCDDTSAGWYGTGQIRACACANDGHCDPTTGRSCAPGNTGLSCQ	861		
QY	137	SACDGDHMGPHCTSRCCCKNG-ALCNPI TGACHCAAGFRGRCEDRCOGTYGNDCHQRC	195		
DB	862	RACDSGHMGPDCTHPCNCSAGHNCDAVSLGCLCEAGYEGPRGQSCRGTYGSCQKC	921		
QY	196	QCONGATCDHYTGRCRCRPGYTGAFCD-----LC 225			
DB	922	RCEGACDHYSGACTCPACWRGSCFCHACPAFFGIDCDACNCSAGACDAVTSGLCIC	981		
QY	226	PRGHGPGQEQRCR-----CONGVCHNTGSCPSGSGMGTVGCPCEGR	272		
DB	982	PAGRWGRCAOSCPPLFTGLNGISQICTCFMGASDSYTGQCHCAPGMWGPCTCDACPPGL	1041		
QY	273	FGKNCSECCCHNGTCDATATGQCCHSCSPGYTGRCODECEVGTGVLCAETCCQVNGGKC	332		
DB	1042	YGRNQCOSCLCRNGRCRDPILGQCTCBEGWTGLACENBCLPRGHNAAGCQNCSTHNGIC	1101		
QY	333	YHVSAGACLEAGFAGECEARLCPGLYGIKCDRCPCPLENTHSCHPMSGECACKPGWS	392		
DB	1102	DLRLGHCLCPAGWTGDKQSS-CVSGTFGVNCEHNCAC--RKASCHHTVAGACFCPPGWR	1158		
QY	393	GLVCNCTSGFGEAGCQICSCQNGADCSYNGKCTCARGFGIDSTPCPTGTGICNG	452		
DB	1159	GPPEQACPRGWFGEAQRCLCPPTNASHVTEGECRCPGFTGLSEQACQPTETEKDC	1218		
QY	453	SSRQGCXKDA-VCSPYDGSCTCKAGWGVDCSIRCPSGTGWGFCGNLTCQCLNGAGCNTLD	511		

Db 1219 EHLCCPPEGTWACDPAAGVCTCAAGHGTGCLQRCPSGRTGCEHCICLNGGTCDDPAT 1278
Qy 512 GTCTCAPGRGECGELPCDDGTGGLNCAERCDSCSHADGCHPTTGHCRCPLPGMSVYCDSDV 571
Db 1279 GACTCPAGAGLADCSLACQGRFGRFSCAHVACACROGACDPVSGACICSPGRTGVRCENG 1338
Qy 572 CAEGRMGPNCSLPC 585
Db 1339 CPDRFRGKCELC 1352

RESULT 2

T27283

hypothetical protein Y64G10A.f - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27283

R:Almscough, R.

submitted to the EMBL Data Library, September 1999

A:Reference number: 220336

A:Accession: T27283

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1620 <WIL>

A:Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB54471.1; CESP:Y64G10A.f

A:Experimental source: clone Y64G10A

C:Genetics:

A:Gene: CESP:Y64G10A.f

A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match

Best Local Similarity 36.7%; Score 1322; DB 2; Length 1620;

Matches 235; Conservative 76; Mismatches 234; Indels 102; Gaps 12;

Qy 16 LCHWT-GRASPLNLEDPNVCSHM-----ESYVTVQESYRPHRPGQIYITGCTDL 64
Db 872 VCHHTVGTCTCLPGKTGRLCDQCLIFETIEFDIAFISIVACAPRTYGPNCATGCS-CV 930
Qy 65 NMFKTRHRVSYRTAYRHGKTMVRRKSCQCPFEYSEGMVPHCAD-----KC 113
Db 931 NCAKDESDGS-----CHCTPGFY--GATGSEVCPTRFGRIDQMLCKC 972
Qy 114 VNGR-CIAPN-TCQCEPRGKGTNCSSACDGHMGRHCTSRCCCKGALCNPTTGACHCAA 171
Db 973 ONGALCDTNSGSECAPGSGRKKCKACAPRTFGKDCSKKCDACADMHCDPSDGEICPP 1032
Qy 172 GFRGRCEDECRCEGRTYGNDCHQRCOCNGAT----- 202
Db 1033 GKKGKCDTCDGLFGAGCKGICSCONGATCDVYTGSECRPGMRGKCDRPPCDGRFG 1092
Qy 203 -----CDHYTGECRCPPGYTGAFCEBDLCPPGKHGPOCEORC 238
Db 1093 EGCNAICDCTTNDTSMVNPFAVARDHYTGECRCAPAGMTGPRCQISCPGLRGRBECRHS 1152
Qy 239 PQONGCVCHHTVGECSGSGMGVTCGQPCPRGFRGKNCSEQCCQNHNGSTCDAAATGQCHC 298
Db 1153 QCSNASCDRTVTFCDGSGFGKNCSECEPRGLGMSNMKHCQLCMHGGCKENGGDEEC 1212
Qy 289 SPGYTGRCQDECPVGTGVLCAETGOCVNGGCKYHVSAGACLEAFGEREARLCPRG 358
Db 1213 IDGMTGPELSL---CPFGQFQRNCAQRNCNKGASCDKRTGRCLDPGWSGEHE-RCVSG 1268
Qy 359 LYGIKRCRCPCHLENTSHCPMSGEACAPGMSGLYCNETSPGFYGEACQOICSQNG 418
Db 1269 HYGACCEETCEC--ENMGALCDPISGHCSCQPGMRGKCKRPLKTYFGHCGQSCRCANS 1326
Qy 419 ADCDSVTCKCTCAPGFKGIDSTPCPLGTGYGNCSSRCCGCKNDAYCSPVDSCTCKAGM 478
Db 1327 KSCDHISRCQCPKGYAGHSTELCPDGFGBGSCQKCCGGSNCDIAISGCKFCPKGHS 1386
Qy 479 GVDCSIIRPSSGTWFGCNLTGCLNGAGACNTLDGTCTCAPRGREGKCELPCQDGTGYGLNC 538

Db 1387 GSDCKSGCVQGRFRPDCNQCLSCENGVCDSSTGSCVCPGRTGTCETACQSDRFGPTC 1446
Qy 539 AERCDSCSHADGCHPTTGHCRCPLPGMSVYCDSDVCAEGRMGPNCSLPC 585
Db 1447 EHCNCENGTCDDRLTGCCRCPLPGFTGHTCNCVCEBGRFAGACKKRC 1493

RESULT 3

T26972

hypothetical protein Y47H9C.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000

C:Accession: T26972

R:Haris, B.

submitted to the EMBL Data Library, October 1998

A:Reference number: 220293

A:Accession: T26972

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1111 <WIL>

A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4

A:Experimental source: clone Y47H9C

C:Genetics:

A:Gene: CESP:Y47H9C.4

A:Map position: 1

A:Introns: 50/2; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1

C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homo1

Query Match

Best Local Similarity 35.7%; Score 1284.5; DB 2; Length 1111;

Matches 246; Conservative 77; Mismatches 221; Indels 175; Gaps 20;

Qy 21 GTASPLNLEDPNVCSHMESYVTVQESYRPHRPGQIYITGCTDLNMFCTBHR 73
Db 35 GTEP---QGDHVC-----VKTIVDY--ELKVIHYVYNDTQCINPLRFGPC----- 80
Qy 74 VSYRTAHRGKTMVRRK-----SQCCEFTYSGE-MCVPHCADKCVHRCIAPNTC 124
Db 81 ---TVERGQKASYOROLVKKREKVKCCDYYOTKHFCLPDCNPCKKCKCEPRGK 136
Qy 125 OCEPGMGNTSSACDGHMGRHCTSRCCCKGALCNPTTGACHCAAGRGMRCE----- 179
Db 137 ECDPEYGGKYCASSCSVGTGWLGCSSCDCEGANCDPELGTCTSGFQGERCKRPPD 196
Qy 180 -----DRCEGRTYGNDCHQRCOCNGAT 202
Db 197 NKMGPNVCYSCPCQNGKCKGKCVSDMGWGERCLMKCEGKRGACCKEPCNQCNGAT 256
Qy 203 CDHYTGECRCPPGYTGAFCEBDLCPPGKHGPOCEORCPONGCVCHHTVGECSGSGMGMT 262
Db 257 CDNTGKCIKISGYHGALENECSVGFPGSGCTQKCDCLNMONCDSSSGECKC-IGMTGK 315
Qy 263 VCGQCPRGREGKNCSEOCOC-----HNNGGCDAAATGOCCHSCPGYTGECROD-ECPTGT 315
Db 316 HCDTCSNRGRELQCKQKNTCGLFESDSNASCDKKTGACCOCESGTYKRPKCDERKCAEQ 375
Qy 316 YGVLCATGTCOCY--NGGKCYHVSAGACLEAFGEREARLCPRGYIKDKRCPCHLE 373
Db 376 YGADSKTCTCVRENTLMCAAPRTGCRCKPFGYGNCEL-ACSRSYSPNCEKQAMCQMN 434
Qy 374 NTHSCHPMSEBACAPGMSGLYCNETGSPGFYGEAKQOICSC-QNGADCDVYTGKCTCAP 432
Db 435 HASECNPETGSCVCPRGRTGKNCSEPCPLDFGPNCALHCCOCNONGVCCDADGCKCQCDR 494
Qy 433 GFKGIDSTPCPLGTGYGNCSSRCCGCKNDAYCSPVDSCTCKAGHNGVDSIRCSSGTMG 492
Db 495 GWTGRCRCHDPRADTFEGANCEKRCCKPRGIGDPTTGCTCPAGLQGANCDIGCEBSYTG 554
Qy 493 FGCNLTGCLNGAGACNTLDGTCTCAPGN-----RGEKCEL--PCQD----- 531
Db 555 PGCKLHCKCVN-GKCDKETGECTQPGFFGSDCSYTTCKSKGYGSECELSRCSDSASCKQ 613
Qy 532 -----GTYGLNCAERCD----- 543

Oy 182 CEQGYGNDCHQRCQCCGATC-DHYTG-ECRCPPGTGACEDLCPPGKKHPOCEORCP 239
 Db 790 I-----NECASN-PCLNQSGCIDVAVGFKCNMLPYTGVECEVTLAP-----CSPR-P 835
 Oy 240 CONGVCHH-----VTGECSPSGMWTGCGPCEPGRFGKNCOSBC-----OCHNGSTCPAA 292
 Db 836 CKNGVCHESDPSFSCNPGAGNOGTCEDV-----NECVANPCTNGVCENTL 885
 Oy 293 TG--QCHCSPGYTGEBRCQ--DECPVGTGYVLAETCCVNGRCY-HVSG-ACLCENAGF 345
 Db 886 RGGGQCRCRNPFGTALCENDIDDC-----EPNCSNGVCQDRVNGFVCCLAGF 935
 Oy 346 AGERCEARL-----CPEGXGKCDKRCRCHLEHNSCHP- 380
 Db 936 RGERCAEDIDCVSAPCNGNCTDCVNSYTCSCPAGESSGINCETNPDCTES--SCFNG 993
 Oy 381 -----MSGECACKPGMSGLYC-----NE-----TCSPEFYGEA 408
 Db 994 GTCVDGJSSFSVCCLPFTGTGYCQHDVNECDSPCQNGSCQDGTGYTKCTCPHGYTGLN 1053
 Oy 409 CQOI---CS---CONGADC--DSYTGKCTCAPFRKIDCSTP----- 442
 Db 1054 CQSLVRWCDSPPCKNGSCWQOGASFCCOCASGWTGICYCDVPVSCEVAARQGVSAVL 1113
 Oy 443 -----CPLGTGYNCSRCG-----CKNDAYCSPVDS--SCTCRAGW 477
 Db 1114 CRHAGCQVDAGNTHLCRCQAGTSTYCOEYDDEQAPNPGANGATIDYLGSTCECVPGY 1173
 Oy 478 HGVDCS-----IRCPSTGTFGCMU---TC----- 499
 Db 1174 HGMCSEINECLSQPCQNGSTCIDLVNTYKSCPRGTQGVHCEIDIDCSPSYDPLTGE 1233
 Oy 500 -OCLNGGACNTLDG--TCTCAPGMRGEKE-----LRQ-DGTGLNCAE-----RC 542
 Db 1234 PRGNGRCVDRVGGYCVCPAGFVGERCEGVNECLSDPDPGSGY--NCVOLINDFR 1291
 Oy 543 DCSHA-----DCCHPT-----TGH-----CRCLPGMSGVHCD----- 569
 Db 1292 ECRFGYTGKRCETVFNCKXDPCKNGGTCAVASTKHGYICKQPGYSGSSCEYDSQSG 1351
 Oy 570 -----SVCAEGRMGPC 581
 Db 1352 SLRCRNGATCVSGHLSPRC 1370

RESULT 6

notch3 protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jul-1998 #sequence_rev:10n 24-Jul-1998 #text_change 02-Aug-2002

C:Accession: S78549; S71825

R:Journal: A.; Tournier-Lasserre, E.
submitted to the EMBL Data Library, April 1997

A:Reference number: S78549

A:Accession: S78549

A:Molecule type: mRNA

A:Residues: 1-2321 <JOU1>

A:Cross-references: EMBL:U97669; NID:q2668591; PIDN:AB91371.1; PID:q2668592

R:Journal: A.; Corpechot, C.; Ducros, A.; Vahedi, K.; Chadrillac, H.; Mouton, P.; Alamowltz

x, M.M.; Weissenbach, J.; Bach, J.F.; Bousser, M.G.; Tournier-Lasserre, E.

A:Title: Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke

A:Reference number: S71825; MUID:97032728; PMID:8878478

A:Accession: S71825

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 67-113;138-194;268-333;G',335-346;536-613;716-765;1240-1279;1815-1888 <JOU2

A:Cross-references: EMBL:U97669

C:Genetics:

A:Gene: notch3

A:Map position: 19p13.1

C:Function:

A:Description: may be involved in pathogenesis of CADASIL, causing a type of stroke and

C:Superfamily: notch protein; ankyrin repeat homology; EGF homology; laminin-type EGF.
 C:Keywords: tandem repeat; transmembrane protein
 F:123-155/Domain: EGF homology <EGX1>
 F:162-194/Domain: EGF homology <EGX1>
 F:240-271/Domain: EGF homology <EGX1>
 F:318-349/Domain: EGF homology <EGX2>
 F:473-504/Domain: EGF homology <EGX3>
 F:853-884/Domain: EGF homology <EGX3>
 F:928-959/Domain: EGF homology <EGX4>
 F:1070-1126/Domain: laminin-type EGF-like homology <LEGG>
 F:1838-1870/Domain: ankyrin repeat homology <AN1>
 F:1871-1903/Domain: ankyrin repeat homology <AN2>
 F:1905-1937/Domain: ankyrin repeat homology <AN3>
 F:1938-1970/Domain: ankyrin repeat homology <AN4>
 F:1971-2003/Domain: ankyrin repeat homology <AN5>

Query Match 19.4% Score 697; DB 2; Length 2321;
 Best Local Similarity 25.28; Pred. No. 1.8e-31;
 Matches 226; Conservative 61; Mismatches 249; Indels 360; Gaps 51;

Oy 5 LNSCLIS-----FICLLCHMIGTASPLNLED-----PNVC-SHMEY 40
 Db 432 VNECLSGPCRNQATCLBRIGQFTICMAGFTGYCEVDIDECQSSPCVNGVCCKDRNGF 491
 Oy 41 SVYQESYPPHFDQIYTTSC--IDILMFKCTRRHVSYPAYRHGEXTMTRRSQCCPGF 98
 Db 492 SCTCPSEFSGSTCQLDVDECASTPCRNAGKVDQPDQY-----ECRCAEGF 537
 Oy 99 YESGEMC---VPHCA-DKCVHRC---IAPTCQCEGMSGTNCSSACDDHMGPHCTSR 151
 Db 538 --EETLDNRVNDSCSPRCHHGRCVVDIASSCAAGYGTRESYVD-----ERSQ 589
 Oy 152 CQCKNGALCNPTTG--ACHCAEGRWRC--DRCEQ--TYG--NDCHQR--CQCONG 200
 Db 590 -PCRHGKCLIDVAKYLRCRPSGTGVNCEVNIIDCASNPCTFCVCHDGINRYCVCYQPG 648
 Oy 201 AT---CDHYTGECRCPPGYTGARCED-----LCPRKGRPOC---EQRC--PCQNG 244
 Db 649 FTGPLCAVEINECASSPCGSGGSCVDENGFRCLCPGSLPLCLPSHPHCAHEPCSH-G 707
 Oy 245 VCHHVTG--ECSCPSGMGTVCQO-----PCEGRFGK 275
 Db 708 ICYDAPGFRVCYCEPBGSGRCSSSLARDACESQPCAGTSSDNGFHTCTPPYQGR 767
 Oy 276 NCS--QEC---QCHNGSTCAATGQ---CHCSPGYTGRCQ---DEC---PVGTYVLC 320
 Db 768 QCELLSPCTPNPCHEHGGRCESAPQPLVPCSPQMGQPRCQDVDECAQAPACGPHCI-C 826
 Oy 321 AE-----TCQ-----CVNKGKCHVSG--ACLCRAGFRAGECEA 352
 Db 827 TNLAGSFCTCHGGYTGSPCDQDINDCPNCLNGGSCQDGVGSFSCSLPGFAGPRC-A 885
 Oy 353 R-----SGBACRPGMSGLYCNE-----LCPEGLYGKICDKRCPCHLEHNSCHPM----- 381
 Db 886 RDVDECLSNPCGPGTCTDHYASFTCTCPPGYGFHCEQDLP-----DCSPSSCFNGGT 938
 Oy 382 -----SGBACRPGMSGLYCNE-----LCPEGLYGKICDKRCPCHLEHNSCHPM----- 381
 Db 939 CVDGVNSFSLCRPGYTGACQHEADPCSLRPLHGGVCSAANPGRCTLESFTGPOCQ 998
 Oy 411 QI---CS---CONGADDSVTGKCTCAPFRKIDC--STP----- 442
 Db 999 TLVDKMSRQPCQNGRCVGTGAYCLCPGMSGRCLDIRSLPCBAAAIQIVRLDQLCQAG 1058
 Oy 443 -----CPLGTGYNCSRCG-----CKNDAYCSPVDS--CTCRAGHGV 481
 Db 1059 GQCVDDESHYCYCPGPRGSHCEGEVDPLAQPCHGGRGCRMYGMYMCECLPYNGDN 1118
 Oy 482 CS-----IRCPSTGTFGCMU---C-----OCLN 503
 Db 1119 CEDDVDECAQPCQHGSCIDLVARYLCSCPSTGLVLEINEDDCGPGPLDSCGRCLH 1178
 Oy 504 GACNNTLDG--TCTCAPGMRGEKEL--PCQDGTGLNCAEKCDSHDCGCHPTTG--- 555

Db 1179 NGCTVDLGGFRCCTCPGTYTGICBADINECRSGA-----CHAATRTCLADDPGGGF 1230
 Oy 556 HCRCLPGMSVHCDY-----CAEGRMGPC 581
 Db 1231 RCLCHAGFSRPGCTVLSPECSPQCHGCRPSPGGLTFTCHCAOPFTWGPC 1286

RESULT 7
 A46019
 Notch-1 protein - mouse
 N:Alternate names: notch protein
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
 C:Accession: A46019; S25144
 R:del Amo, F.; Gendron-Maguire, M.; Swiatek, P.J.; Jenkins, N.A.; Copeland, N.G.; Grid
 Genomics 15, 259-264, 1993
 A:Title: Cloning, analysis, and chromosomal localization of Notch-1, a mouse homolog of
 A:Reference number: A46019; MUID:93194170; PMID:8449489
 A:Accession: A46019
 A:Status: not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 1-2531
 A:Cross-references: GB:211886; GB:S47228; NID:q288502; PIDN:CAA77941.1; PID:q288503
 A:Note: sequence extracted from NCBI backbone (NCBI:127318)
 R:Francisco del Amo, F.; Smith, D.E.; Swiatek, P.J.; Gendron-Maguire, M.; Greenspan, R.J.;
 submitted to the EMBL Data Library, April 1992
 A:Description: Expression pattern of Notch, a mouse homolog of Drosophila Notch, suggest
 A:Reference number: S25144
 A:Accession: S25144
 A:Molecule type: mRNA
 A:Residues: 1551-2108, 'Q', 2110-2114, 'ALP', 2118-2170 <FRA>
 A:Cross-references: EMBL:Z11886
 C:Genetics:
 A:Gene: notch-1
 A:Map position: 2
 A:Note: proximal region of chromosome 2
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:106-18/Domain: EGF homology <EGF1>
 F:144-175/Domain: EGF homology <EGF1>
 F:222-254/Domain: EGF homology <EGF2>
 F:261-282/Domain: EGF homology <EGF2>
 F:339-370/Domain: EGF homology <EGF3>
 F:416-449/Domain: EGF homology <EGF3>
 F:456-487/Domain: EGF homology <EGF4>
 F:494-525/Domain: EGF homology <EGF5>
 F:533-563/Domain: EGF homology <EGF6>
 F:607-638/Domain: EGF homology <EGF7>
 F:682-713/Domain: EGF homology <EGF8>
 F:757-788/Domain: EGF homology <EGF9>
 F:795-826/Domain: EGF homology <EGF10>
 F:873-904/Domain: EGF homology <EGF11>
 F:911-942/Domain: EGF homology <EGF12>
 F:949-980/Domain: EGF homology <EGF13>
 F:987-1018/Domain: EGF homology <EGF14>
 F:1025-1056/Domain: EGF homology <EGF15>
 F:1063-1094/Domain: EGF homology <EGF16>
 F:1149-1180/Domain: EGF homology <EGF17>
 F:1187-1218/Domain: EGF homology <EGF18>
 F:1233-1264/Domain: EGF homology <EGF18>
 F:1352-1383/Domain: EGF homology <EGF19>
 F:1391-1425/Domain: EGF homology <EGF>
 F:1917-1948/Domain: ankyrin repeat homology <AN1>
 F:1949-1981/Domain: ankyrin repeat homology <AN2>
 F:1983-2015/Domain: ankyrin repeat homology <AN3>
 F:2016-2048/Domain: ankyrin repeat homology <AN4>
 F:2049-2081/Domain: ankyrin repeat homology <AN5>

Query Match 19-2*, Score 693; DB 2; Length 2531;
 Best Local Similarity 25-3*, Pred. No. 3-2e-31;
 Matches 217; Conservative 71; Mismatches 206; Indels 364; Gaps 51;
 Oy 10 SFICLLCHWIGTASPLNLEDPNVCWSHWSYSVTQESYPPHPDQIYTTCTDILNMFRC 69

Db 624 SYLCCLGKGTGPNCENILDD---CA-----SNPCDS---GTCIDKIDGEC 664
 Oy 70 TRHRVYTAIRRHGKTYRRKRSQCCPGFYEGEMC---VPRCA----- 110
 Db 665 A-----CEPGY---TGSNCNVNIDECASPCPNHNGTECDGIAG 699
 Oy 111 -----DKVHGRC---IAPNTCOCEPMGGTNC---SSACDG 141
 Db 700 FTGCRPEGYHPTCLSEVNECNSNFCIHGACDGLNGKCDCAPEMSGINDINNDECS 759
 Oy 142 DHMGPHCTSRQCKNGALCNPTTG---ACHCAFGFGRNCEDECGTYGNDCHORCOQON 199
 Db 760 N-----PCVNGGCKDKMTSGVYCTCREGFGSPNCO-----TINIECASN-PCIN 802
 Oy 200 GATC-DHYTG-ECRCPPTGTAFCDLCPPGKHGQCRCORCPQONGVC---HHVTGEC 253
 Db 803 GGTCTIDVAGYKCNCPDLYTGATCEVYLA-----C-ATSPCKNSGVCKESEDYEFSC 855
 Oy 254 SCPSGMGNTVC-----GQPCPEGR-----FGKNCN---QRCQ--- 282
 Db 856 VCPFGMOGTCEVDINECVKSPCRHGAASCQNTNGSYRLCLQAGTYGRNCESDIDCRNP 915
 Oy 283 CHNGGTCDAA---TGQCHSPGYTGERCODE-----CPVGT 315
 Db 916 CHNGSGCTDGIATFACDCLPGHGAFCEDINECASNPQONGANCCTDQVDSYTCPIYGF 975
 Oy 316 YGVLCAEI-----CQCVNGKCIYVSG---ACICEGAFAGEAREARLCPGLYGT-KC 364
 Db 976 NGIHENNTPDCTESSCENGTC---VDGINSTCLCPPTGSGSYCO-----YDVNEC 1025
 Oy 365 DKRCCHLENTHSCHPMNSG---ECACKPGMSGLYCNE-----TCSPGFYGEACQOI----- 412
 Db 1026 DSR-PCLHGGT---CODSYGTYKCTCPQGTGTGINCONLYRMCDASAPCKNGRGMQNTQYH 1082
 Oy 413 CSCQN---GADCDVYTGKCTCAPGRKIDCSTPCPLGTYGIN-----CSSRQ----- 457
 Db 1083 CECRSRGWGVNCDVLSVCEVAQKRGIDVTLQCHGGGLCVDEGDKHYCHCQAGYTGSYC 1142
 Oy 458 -----CKNDVCSPPVDG---SCITCKAGMHGVDS----- 483
 Db 1143 EDEVDKSPNPNQNGATCTDYLGGFSCKCVAGYHNSNSEELINECLSQPCONGTCTDILT 1202
 Oy 484 ---TRPSGTWFGCENLT---C-----OCLNGACNTLDG---TCTCAPGRME 523
 Db 1203 NSYKSCPRGTGVCHEINVDCHPPLDPASRPFNNGTCVDYVGGTCTCPGPFVGE 1262
 Oy 524 KCE-----LPCOD-GTYGLNCAER-----SHADG----- 550
 Db 1263 RCEGVNCLSNPCDPRGTQ---NCVQRVNDFHCRCRAGHTGRCSYVINGCRKPKCKNG 1320
 Oy 551 -----HPTTGH-CKCLGMSGVHDS-----VCACGRGPNP----- 581
 Db 1321 VCAVASNTARGFTCRCPGFEATCENDARTCGSLRCLNGTCTIGSPSPCTCLGSPFTG 1380
 Oy 582 -----SLPCY 586
 Db 1381 PECOPASSPCVGSNPCI 1398

RESULT 8
 A49175
 Notch B protein - mouse (fragment)
 N:Alternate names: Notch homolog
 C:Species: Mus musculus (house mouse)
 C>Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
 C:Accession: A49175; PHL570; S32113
 R:Iardelli, M.; Lendahl, U.
 Exp. Cell Res. 204, 364-372, 1993
 A:Title: Notch A and Notch B--two mouse Notch homologues coexpressed in a wide varie
 A:Reference number: A49175; MUID:93178563; PMID:8440332
 A:Accession: A49175
 A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA
 A:Residues: 1-1203 <LARI>
 A:Cross-references: EMBL:668279; NID:g287989; PID:CAA8340.1; PID:g287990
 A:Experimental source: embryo
 A:Note: sequence extracted from NCBI backbone (NCBI:126158)
 C:Comment: This protein has many EGF repeats and 11n-12/Notch repeats.
 C:Comment: This protein is one of the neurogenic proteins controlling the decision between
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:143-174/Domain: EGF homology <EGF1>
 F:482-513/Domain: EGF homology <EGF1>
 F:560-591/Domain: EGF homology <EGF1>
 F:674-705/Domain: EGF homology <EGF2>
 F:712-743/Domain: EGF homology <EGF3>
 F:836-867/Domain: EGF homology <EGF3>

Query Match 19.1%; Score 687; DB 2; Length 1203;
 Best Local Similarity 24.8%; Pred. No. 4,3e-31;
 Matches 221; Conservative 78; Mismatches 245; Indels 348; Gaps 56;

```

3 ISLNSLSFLCT---LICH-----WIGTASPLNLE--DPNVCSSM-----ES 39
214 IDIDCSSTPLNGAKCIDHPNGYECQCATGFTGILCDENIDNCDDPCHHGQCCODIDS 273
40 YSTVQESYPRHF--DOI--YRYS-----CTDILNMEKCT-----RHVSY---- 76
274 YTCINPMTWGAICSDQIDECYSSPCLDNGRCIDLNVGYQCNCQPGTSLNCEINDDCA 333
77 RTAYRHS--EKTMYRRKSQCCPGF-----YESGEMCV----- 106
334 SNNCHGVGVVDGINRYSVCSPFTGQRCNIDIDECASNCRKAGATCINDVNFRCICPE 393
107 ----PHC-----ADKCVHGRN---IAPNTQCEPBGWGTNCSSACDGHMGPCTSR 151
394 GRHPHPCYSOVNBECLNPNCHGNCTGLSGYKCLCDAGWGVNCE--VDKN---ECLSN 447
152 COCKNALCNPIIGA--CHCAAFGRMCE--DRC-----EGTYGNCH--QRQCCO-- 198
448 -PCQNGCTNNLVNRYRCKTKKGFKNQCVNIDECASNCLNCGTCFDDVSGYTHCML 506
199 --NGATCDHYTGRCRPRGYTGAFCED-----LCPRKHGPOCE--QRC--PCQ 241
507 PTGKNCQYTLAPCSNPENCAVKEARPFESFSLCAPMGKRCQTYVDCEISKPCM 566
242 NGCVCHHVTGE--CSCPBGMTVCQPCPEGRFKNCSQEC--OCHNGTIC--DAATG 294
567 NNGVCHNTGSSYVCECPRGSGMDCEDI-----NDCLANPCONGSCVDHYNTR 616
295 OCHCSGYTGERCODE-----CPVGYTYGVC--AETC-- 324
617 SCQCHPFTGDKCQTDMECLSEPCKNNGTCSDYVNSYTCCTPAFGHGVHCENNIDECTE 676
325 -OCVNGKCYHVSQ---ACLCBAGFAG-----ERCEAR-----LC 355
677 SSGFNGTIC--VDGINSFGLCPVFTGPRCLHDINECSSNPLNAGTICVDGLTYRCIC 734
356 PEGLYGICKD-----KRCPCLENTHSCHPMSEGCACRPGMSGLYCNE----- 398
735 PLATYTKNCOITLVNLCRSRCKKNGTCVDEKARPHCLCPRGMDGAYCDVLVNSCAALQ 794
399 -----TSPGFYGAACQ-----ICS-----CQNGACDDSYTG-- 426
795 KGVVPHLCOHSGICINAGNTNHCQCPLEGTSYCEQULDEECASNPCOMHATNDNFIQY 854
427 KCTCAFGKIGIDSTPCPLGTGYGINSRQCKNDVAVCSYDSSCTCKAGMHVDSIRL 486
855 RECCEVGYGVNCE-----YEVDCQNPONGTCTIDLNVHFKCS-----C 896
487 PGTWFGG--NL--TC-----QCLNGAGC--NTLDG--TCTCAPGWRGKCE-----LPC 529
897 PGTRELLCEBNDIDECAGGPHCLNGGQVDRIGYTCRCCLPFAAGRCBEDINECLSNPC 956
530 -ODGTYGLNCAE-----RCDCHSA-----DGC-----HPTTGIC 557

```

Db 957 S5EGS--LDCVOLKNYNYNCICRSATFTGRHCEFLDYCPQKPCNLNGGTCAVANSMPDGLIC 1014
 QY 558 RCLPGMSGVCHDSVCAEGRW-----GPNC-----SLPC 585
 Db 1015 RCPGFGSGARCCSSCGQVCKRGEDQCIHDSGFRICLPKDCESGCASNPC 1066

RESULT 9

notch 3 protein - mouse
 S45306

C:Species: Mus musculus (house mouse)
 C:Date: 20-Feb-1995 #sequence, revision 20-Feb-1995 #text, change 02-Aug-2002
 C:Accession: S45306

R:Lardelli, M.; Dahlstrand, J.; Lendahl, U.
 Mech. Dev. 46, 123-136, 1994

A:Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth fact
 A:Reference number: S45306; MID:95001556; PMID:7918097

A:Accession: S45306
 A:Status: preliminary

A:Molecule type: mRNA
 A:Residues: 1-2318 <LARI>

A:Cross-references: EMBL:74760; NID:g483580; PID:CAA52776.1; PID:g483581
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:163-195/Domain: EGF homology <EGF1>
 F:474-505/Domain: EGF homology <EGF1>
 F:854-885/Domain: EGF homology <EGF2>
 F:1839-1871/Domain: ankyrin repeat homology <AN1>
 F:1872-1904/Domain: ankyrin repeat homology <AN2>
 F:1906-1938/Domain: ankyrin repeat homology <AN3>
 F:1939-1971/Domain: ankyrin repeat homology <AN4>
 F:1972-2004/Domain: ankyrin repeat homology <AN5>

Query Match 19.0%; Score 685.5; DB 2; Length 2318;
 Best Local Similarity 24.4%; Pred. No. 7.9e-31;
 Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

```

94 CCEGYESEGMEVPHCADKCVHGRCTAPNT-----CQCEPBGWGTNCSSACDGHW 144
225 LIPGF--EGQNGCEVN--VDPCRGHRCINGGTGCVGVNTYVNCQCPPEWTGQCTEDVD-- 277
145 GRHCTSRQY-----CKNGALCNPIIG--ACHCAAG-----FRGWRCE 179
278 -----ECQLQPNACHNGGTCFNLIDGHSVCVNGMTGESCSQNDIDCATVCFHGATCH 331
180 DR-----CEQTYGNDCN--ORC--OCONATCD--HYTGE--CRCPGYTGAFCED 223
332 DRVASFYCACPMKGTGLCHLDACVSNPCHEDIAIDTNVSRALCTQPRPTGGACDQ 391
224 -----LCPRGHGRPOCE----- 235
392 DVDECSIGANPCBHLGRCVNTQGSFLQCQCGRGYTGPRCETDVNECLSGPCRNQATCIDRI 451
236 -----ORCPONGGVC--HHVYG--ECSCPGMMGTVC----- 264
452 GQFTCIOMAGFTGYEVDIDECQSSPCVANGVCKDRVNEFSCTCPSGSGSMCOLDVDE 511
265 -----GQP-----CPEGRFKNCSQ--EQ--CHNGGTCDA--ATGQCIC 298
512 CASTPCRNKAKCVDPDGYEBCRCABFEGTLCERANNDDCSPPDCHHGRVYGLASSCAC 571
299 SPGYTGERCODE-----CPVGYTYGVC-----AETCQ 325
572 APGYTGRCSGVDECRSQPCRYGKCLDLVDYKLCRCPGTGTVNCEVNIIDCASNPCY 631
326 ---CVNGKCYHVSQCLCEAGAGERCARL-----CPEGLYGIKCDKRCR 369
632 FGVCRDGINRYD---CYCQPGFTGPLCANEINEGASSPCGEGSGCVDENGEHC--LCP 665
370 -----CHLEHTHS--CHPMSSG--ECACKPMSGGLYCNE----- 398
686 PGLSLPLCLPANHPCAKHRCSSHVCHDABGFGFCVCEPFGHSGRCQSLAPDACESQPCQ 745
399 -----TSPGFYGAACQOI--CS--CQNGACDDSYTGK--CTCABGFKG 436

```

[illegible]

```

RESULT    10
A49128     cell-fate determining gene Notch2 protein - rat
C:/Species: Rattus norvegicus (Norway rat)
C:/Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 02-Aug-2002
C:/Accession: A49128
R:/Weinmaster, C.; Roberts, V.J.; Lemke, G.
Development 116, 931-941, 1992
A:/Title: Notch2: a second mammalian Notch gene.
A:/Reference number: A49128; MUID:93202015; PMID:1295745
A:/Accession: A49128
A:/Status: preliminary; not compared with conceptual translation
A:/Molecule type: mRNA
A:/Residues: 1-2471 <WEI>
A:/Experimental source: Schwann cell
A:/Note: sequence extracted from NCBI Backbone (NCBIF:127811)
C:/Superfamily: notch protein; ankrylin repeat homology; EGF homology
F:/264-295/Domain: EGF homology <EGX1>
F:/799-830/Domain: EGF homology <EGT1>
F:/877-908/Domain: EGF homology <EGX2>
F:/1029-1060/Domain: EGF homology <EGF>
F:/1067-1098/Domain: EGF homology <EGX3>
F:/1153-1184/Domain: EGF homology <EGFX>
F:/1191-1222/Domain: EGF homology <EGK4>
F:/1876-1908/Domain: ankrylin repeat homology <AN1>
F:/1909-1941/Domain: ankrylin repeat homology <AN2>
F:/1943-1975/Domain: ankrylin repeat homology <AN3>
F:/1976-2008/Domain: ankrylin repeat homology <AN4>
F:/2009-2041/Domain: ankrylin repeat homology <AN5>

Query Match      19.0%; Score 685.5; DB 2; Length 2471;
Best Local Similarity   24.8%; Pred. No. 8.2e-31;
Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55

OY          3 ISLNSCLSFICL-----LLCH-----WIGRASPLNLF--DPNVCSHW-----ES 39
               ||| | | | | | | | | | | | | | | | :|
Db         531 IDIDCSSPPTSGAKACIDHPNGHYECOCATGTTGLTCENINDCPDPCDHGCCOPDGIDS 590

OY          40 YSVTOEASYPHF--DOI--YYTS-----CTDLINMEKF----RRRVSY----- 76
               ||| | | | | | | | | | | | | | | | :|
Db         591 YTGTGNPGTYMGALGSDQIDDECSSSFLCDNDGRCIDLYAGTQCCKPETSGLNELTNDDCA 650

OY          77 RTAYHNGE--KTMYERRKSOCCEPFYESGEHC---VPHCADK----- 112
               || | | | | | | | | | | | | | | | | :|
Db         651 SMLPSHGACVUDGINNYSCVSPPF--TGRCRIDIDECCSNCRKDATCINDYNPRCMC 708

OY          113 -----CVHGMC---IADNTQCEBERGGWGTCSSACDGDHMWRPCT 149
               ||| | | | | | | | | | | | | | | | :|
Db         709 REGPNRPSPQYSUOVNECLSSPRTGHNGCGTGLSGKYSLCLDVAGNGVINCE--VDKN----ETL 762

```

QY	150	SRQCKKALCNPLTGA--CHQAAFGPMGAC---DRC-----DQGTGNDCH-QRQC	197
Db	763	SN-PCQNGGTCTNNLVNGYRCTCKKGFKXGYNQVNDCEASNPCLNGTCLDDBVSGYTC	821
QY	198	Q--NGATCDHYTGECRCRPGYTGAFED-----LCPRGKHGPOE--ORC---P	239
Db	822	MLPTGKNCQVTLARCPSPNCEMAAYCKEAPNESTCLCARQMGQGRCTVYDVCYKAP	881
QY	240	CQNGVCHHTGTE--GSCPSGMNGTYCGQPCPGRFGKNCQBC--QCHNGTC--DAA	292
Db	882	CMNNGICHTNONGSYMCCECPGFSFGMDCEDDI-----NDCLANPCQNGSCVDVYN	931
QY	293	TGQCHSGPYTGEMCE-----CPVGTGLVC---AENG	324
Db	932	TFSCCLRLPGLFGDDCQTDNLVNLCSBPCKNKGTCSDYVNSTYCTCPAGHGVHCENNIDEC	991
QY	335	--OCVNGGRCYHYSG---ACLCENAGFAGERC-----EARLCEPGYGLK	364
Db	992	TESSCFNGTIC--VDGINSFSCLCVPYGFPTFLHDINECSSNPCLNGSTCVDJGTYRC	1049
QY	365	-----DKRC-----PCHLEHTHSCHPMSEBCKRQWGSGLYNE-----	398
Db	1050	TCPLGTYGKNQOTLVNLCSBPSCKNKGTCQOEKARPRCLCPGMDAYADVLVNSCKAA	1109
QY	399	-----TSBGFYGEACOO---ICS---CONGADDSTYG	426
Db	1110	LQKPYVEHLCOHSGICLIMGNTHHCOCPPLGTYGSCICEBDLDECBANPCQPHATSDPTIG	1169
QY	427	--KCTCAPEFKGIDCSTPCEPLGTYGINCSSRCCCKNDAYCSPVDSGCTCKAGMHGDCSI	484
Db	1170	GYRECBVGYGYNVCE-----YEVEDCONPCQNGGCTCIDLVNHFKCS-----	1212
QY	485	RCPSGTMGFGC--TL-TC-----QCLNGAC-nTLdG-TCTCAPEGMGEC-----L	527
Db	1213	-CPGTKRLCLAEENIDDCAGAPHLCLNGGQCVDRIGYISCRCLPGEFAGECEBDINECLSN	1271
QY	528	PC-ODGTGYLNCAR-----RCDS-SHA-----DGH-----PTTG	555
Db	1272	PCSEBGS--LDCIDLKNNGYOCVCRSAFNTGHCEFTFLDYCPQRKPCULNGICAVNAVPPGF	1329
QY	556	HCRCLPMGSGVHCDVSCABGRW-----GPNCSLP	584
Db	1330	ICRCPGFSGARCSGCGOVKCRGCECVHTAASGPHCFCP	1369

RESULT 11

A40701
tenascin-X precursor - human
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
C:Accession: A40701; A33725; C42175
R:Ristow, J.; Tee, M.K.; Gitelman, S.E.; Mellon, S.H.; Miller, W.L.
J Cell Biol. 122, 265-278, 1993
A:Title: Tenascin-X: a novel extracellular matrix protein encoded by the human XB gene
A:Reference number: A40701; MUID:93300909; PMID:7686164
A:Accession: A40701
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3566 <BRI>
A:Cross-references: EMBL:X71937
R:Morel, Y.; Bristow, J.; Gitelman, S.E.; Miller, W.L.
Proc. Natl. Acad. Sci. U.S.A. 86, 6582-6586, 1989
A:Title: Transcript encoded on the opposite strand of the human steroid 21-hydroxylase
A:Reference number: A33725; MUID:89367293; PMID:2475872
A:Accession: A33725
A:Molecule type: mRNA
A:Residues: 2748-3199, 'V', 3201-3298, 'E', 3299-3314, 'G', 3316-3566 <MOR>
A:Cross-references: GB:M55813; NID:9183069; PIDD:AAA35884.1; PID:9183070
R:Matsumoto, K.; Arai, M.; Ishihara, N.; Ando, A.; Inoko, H.; Ikemura, T.
Genomics 12, 485-491, 1992
A:Title: Cluster of fibronectin type III repeats found in the human major histocomp

A:Reference number: A42175; UID:92217969; PMID:1373119
 A:Accession: C42175
 A:Molecule type: DNA
 A:Residues: 1849-1936 <MAT>
 A:Experimental source: clone 3.9Kf3-1
 A:Note: sequence extracted from NCBI backbone (NCBIP:95694)
 C:Genetics:
 A:Gene: GDB:TNXA; D6S103E; TNX; XA; XB
 A:Cross-references: GDB:568487; OMIM:600261
 A:Map position: 6p21.3-6p21.3
 C:Keywords: extracellular matrix; glycoprotein
 C:Keywords: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty
 F:435-461/Domain: EGF homology <EGF>
 F:748-828/Domain: fibronectin type III repeat homology <3F1>
 F:829-856/Domain: fibronectin type III repeat homology <3F2>
 F:873-953/Domain: fibronectin type III repeat homology <3F3>
 F:975-1055/Domain: fibronectin type III repeat homology <3F4>
 F:1078-1158/Domain: fibronectin type III repeat homology <3F5>
 F:1167-1247/Domain: fibronectin type III repeat homology <3F6>
 F:1248-1317/Domain: fibronectin type III repeat homology <3F8>
 F:1323-1403/Domain: fibronectin type III repeat homology <3F9>
 F:1412-1492/Domain: fibronectin type III repeat homology <3F10>
 F:1510-1590/Domain: fibronectin type III repeat homology <3F11>
 F:1618-1676/Domain: fibronectin type III repeat homology <3F12>
 F:1678-1749/Domain: fibronectin type III repeat homology <3F13>
 F:1751-1831/Domain: fibronectin type III repeat homology <3F14>
 F:1849-1929/Domain: fibronectin type III repeat homology <3F15>
 F:1955-2035/Domain: fibronectin type III repeat homology <3F16>
 F:2061-2141/Domain: fibronectin type III repeat homology <3F17>
 F:2167-2246/Domain: fibronectin type III repeat homology <3F18>
 F:2274-2354/Domain: fibronectin type III repeat homology <3F19>
 F:2382-2462/Domain: fibronectin type III repeat homology <3F20>
 F:2488-2568/Domain: fibronectin type III repeat homology <3F21>
 F:2584-2664/Domain: fibronectin type III repeat homology <3F22>
 F:2671-2757/Domain: fibronectin type III repeat homology <3F23>
 F:2771-2851/Domain: fibronectin type III repeat homology <3F24>
 F:2878-2958/Domain: fibronectin type III repeat homology <3F25>
 F:2977-3067/Domain: fibronectin type III repeat homology <3F26>
 F:3078-3159/Domain: fibronectin type III repeat homology <3F27>
 F:3167-3247/Domain: fibronectin type III repeat homology <3F28>
 F:3255-3334/Domain: fibronectin type III repeat homology <3F29>
 F:3349-3557/Domain: fibrinogen beta/gamma homology <FBG>
 Query Match 18.9%; Score 682; DB 1; Length 3566;
 Best Local Similarity 28.4%; Pred. No. 1.6e-30;
 Matches 191; Conservative 41; Mismatches 178; Indels 262; Gaps 38;

Db 418 RCVCWBYGTYTDCGSRACPRDCRGRCRNGVCVNAAGTSGEDCGVRSPPDCRGRCRCE 477
 Oy 423 SVTKCTCAPGFKGIDCST-----PCPLGTGYNICSS-RC--GCKND 461
 Db 478 S--GRCMCPYGTGRCCGRACRGGRCVGRVGRVGNCPGTGIDCSRRRCRGRBH 535
 Oy 462 AVCSPYDSSCTCAGNHWYDCSTIR-CPSTWFGCNLTQCLNG----- 504
 Db 536 GLCE--DGVACVDAGYSGEDCSTRSCPGGCRGRG-----QCLDGRVCEDGYSGEDCGVR 588
 Oy 505 -----GACNTLDGTCTCAPGRCRGCLEP-----CODGYGLN 537
 Db 589 QCPNDCSQHVCO--DGVCTCWEGYSEDCSITPCSNHGRCRCEGRCLDPGTYTGP 646
 Oy 538 CAER---CDCSHADGCHPTTGHRCCLPGWSGVHC-----DSY 571
 Db 647 CATRMCPACRGRGR--YGVCLCHVYGGEDCGGEEPPASACPGCGCPRELCTRAGQCV 704
 Oy 572 CAEGRMGPMCSL 583
 Db 705 CVEGFRGPDCAI 716
 RESULT 12
 T31070
 notch homolog - sea urchin (Lytechinus variegatus)
 C:Species: Lytechinus variegatus (variegated urchin)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T31070
 R:Sherwood, D.R.; McClay, D.R.
 Development 124, 3363-3374, 1997
 A:Title: Identification and localization of a sea urchin Notch homologue: Insights in
 A:Reference number: EMBL:AF000634; NID:92570350; PID:92570351; PIDN:AA82088.1
 A:Accession: T31070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2531 <SH>
 A:Cross-references: EMBL:AF000634; NID:92570350; PID:92570351; PIDN:AA82088.1
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 Query Match 18.9%; Score 681.5; DB 2; Length 2531;
 Best Local Similarity 24.2%; Pred. No. 1.4e-30;
 Matches 216; Conservative 68; Mismatches 215; Indels 393; Gaps 53;


```

Db 702 SNINESSNPVNGSGCHDVNEYSCECMAGYTGTCTDPEBC-----SSNPGQ 751
Qy 328 NGKGC--YVSGACILCEAFAGEARL-----CPEGLYG 361
Db 752 HGGTCNRAFYNCCTQADYTTGNCENVNIDCDVEPLNGICIDEVNSFOCCVCPQTFVG 811
Qy 362 IKCD-KRCCHLENTSHCPMSECCACKPGMSGLVCNENCSPEFVEACQO-----ICS 414
Db 812 LICETFRSPC---EDNOCQ-NGATCYSEDYAGYSCR--CTSGFOCNFCDDDRNECLFSP 865
Qy 415 CONGADCDSDVTG--KCTCAPGFKGIDC-----STP-----CPLGT 447
Db 866 CRRGSGCTNLGSEFECSCPLGPDGPICEINIDECASGCPCTNGGICITDLIDYFCSQGRF 925
Qy 448 YGINC---SSRC---GCKNDVAVSP-VDG-SCTCKRAGMHGVDCISCPGSGTMCNLTC 499
Db 926 TERNQNDDECLSSPCRNAGATCHEYVDYTCISLGVFGSMHCEINDODCT-----TS 978
Qy 500 OCLNGACNTLDG---TCTCAPGMRGKCELC-----PCDDGTGLNCAER----- 541
Db 979 SCLYGTCTC---IDGVNYSYTECCTGYTGSCNQIIEINCDSDPCENGA---TCDDRFSYS 1033
Qy 542 -CD-----CSH-ADGCHP-----TTGH--CRCLPGMSGVHCD----- 569
Db 1034 HCDVGFGLNCEHYVQWCSPONNPCYNATGVAMGHLIECHCASNMGKLCIDVPKYSQDI 1093
Qy 570 -----SYCAGRMGQPC-----SLPCY 586
Db 1094 AASDKNVTRELCNGTCTIDATSSHSLCQDGYGSGYCEVNIDECASAPCH 1145

```

RESULT 13

A24420

notch protein - fruit fly (*Drosophila melanogaster*)
 M:Alternate names: neurogenic repetitive locus protein
 C:Species: *Drosophila melanogaster*
 C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text change 10-Sep-1999

A:Accession: A24420; A24768; S09358; A05267
 R:Kidd, S.; Kelley, M.R.; Young, M.W.
 Mol. Cell. Biol. 6, 3094-3108, 1986

A:Reference number: A24420; MUID:87064624; PMID:3097517
 A:Accession: A24420
 A:Molecule type: DNA

A:Residues: 1-2703 <KID>
 A:Cross-references: GB:K03508; NID:9157993; PIDN:AA28725.1; PID:9157993

R:Wharton, K.A.; Johansen, K.M.; Xu, T.; Artavanis-Tsakonas, S.
 Cell 43, 567-581, 1985

A:Reference number: A24768; MUID:86079539; PMID:3935325
 A:Accession: A24768
 A:Molecule type: mRNA

A:Residues: 1-48, '1', '50-118, 'R', '120-230, '1', '232-256, 'N', '258-266, 'A', '268-872, 'R', '874-958, 'R', '958-1000, 'D'.
 A>Note: the authors translated the codon ATC for residue 49 as Thr, ATT for residue 2044

Nucleic Acids Res. 17, 6463-6471, 1989
 A:Title: Hypervariability of simple sequences as a general source for polymorphic DNA ma
 A:Reference number: S09358; MUID:89385974; PMID:2780284

A:Accession: S09358
 A:Molecule type: DNA
 A:Residues: 2505-2551, 'Q', '2552-2576, 'E', '2578-2604 <TAU>
 R:Wharton, K.A.; Yedvobnick, B.; Flinnerty, V.G.; Artavanis-Tsakonas, S.
 Cell 40, 55-62, 1985

A:Title: opa: a novel family of transcribed repeats shared by the Notch locus and other
 A:Reference number: A05267; MUID:85099329; PMID:2981631
 A:Accession: A05267

A:Molecule type: DNA
 A:Residues: 2504-2576, 'E', '2578-2611 <WHA2>
 C:Genetics:

A:Gene: notch: opa
 A:Cross-references: FlyBase:FBgn0004647
 A:Map position: 8.96-9.36

A:Introns: 53/3; 84/3; 171/3; 240/3; 283/3; 2333/3; 2436/3; 2588/3
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology

C:Keywords: differentiation; tandem repeat; transmembrane protein
 F:27-43/Domain: transmembrane #status predicted <TM1>
 F:297-328/Domain: EGF homology <EGF1>
 F:530-561/Domain: EGF homology <EGF1>
 F:568-599/Domain: EGF homology <EGF>
 F:988-1019/Domain: EGF homology <EGF2>
 F:1064-1095/Domain: EGF homology <EGF3>
 F:1187-1218/Domain: EGF homology <EGF3>
 F:1746-1762/Domain: transmembrane #status predicted <TM2>
 F:1950-1982/Domain: ankyrin repeat homology <AN1>
 F:1983-2015/Domain: ankyrin repeat homology <AN2>
 F:1988-2004/Domain: transmembrane #status predicted <TM3>
 F:2017-2049/Domain: ankyrin repeat homology <AN3>
 F:2050-2082/Domain: ankyrin repeat homology <AN4>
 F:2083-2115/Domain: ankyrin repeat homology <AN5>
 F:2358-2368/Region: glutamine-rich
 F:2538-2568/Domain: neurogenic repetitive element #status predicted <OPA>

Query Match

18.8%; Score 677; DB 1; Length 2703;

Best local similarity 25.4%; Pred. No. 2.6e-30;

Matches 208; Conservative 78; Mismatches 203; Indels 330; Gaps 51;

```

Qy 7 SCL---SFICLLCHWIGTASPLNED--PNVCSHWESYVTVQESYHPEDQIYYTSC 60
Db 502 SCLDPEGTFRVCMPGFTGTQCEIDIDECQSNPC-----LNDGTC 541
Qy 61 TDLNFKCTRRNRYRTAIRRGEKTYRRKSCCCGFYSESGMC---YPRCADKCYHGR 117
Db 542 HDKINFEKS-----CALGF--TGARCCQINIDDCQSQPCNRN 576
Qy 118 -----CIAPNTCOCPEMGNTGNS--SACDGDHMGPHCTSRCCNKALCNPTG-A 168
Db 577 GICHDSIAGYSCCEPRPGYGTCTCEININDCDSN-----PCIRGACIDVNSFKKL 626
Qy 169 CAAGFRGW-----RCEDR-----CEGTYG-----NDCHQRCQ 196
Db 627 CBPGYGYICOKQOINCEESNPQCFDGHODRVGSYYCQCCAGSRKNCVNECHSN-P 685
Qy 197 CONGATC-DHYNG-ECRCPRGYTGAFCELCPRGKRGPCQEDRCPCONGVC-HYNTG-E 252
Db 686 CNGGATCIDGINSYKQCVPGFTGQHC-----KNVDKIS-SPANNVCIDVONGYK 738
Qy 253 CSCPSGMMGTVC-----GOP-----CEGRGRKNCN--OECQ-- 282
Db 739 CECPRGFYAHCLSDVDEBANSNPNVNEGRCEDEINEFICHPRGYGRKCEIDIDECSSN 798
Qy 283 -CHNGTGC-DAATG-QCHCSPGYTGERCQ--DECPVGYGYVCAETCCQVNGKCY-HV 335
Db 799 PCOHGTCYDKLNAFSCQCPGTYGQKCEFNIDDC-----YTNPCGNGTCTIDV 848
Qy 336 SG-ACICEAGFAGEARCEALCEGLYGINC-DKRCCHLENTSHCHPNSG---ECACKP 389
Db 849 NGIKCYCKAPFYGRDCEKMDP-----CASNRC--KNEAKCTPSSNFDLFCSTCYL 897
Qy 390 GMSGLYCNE-----TCSPGFYGEAC--QOICS--CON 417
Db 898 GTTGRCCDDIDECSSLSPCRNGASCLNPGSYRCLCTGTGEBRDCAINMTDCASPCQN 957
Qy 418 GADCDSDVTG--KCTCAPGFKGIDCST-----PCPLCTYGI 450
Db 958 GGTCTDGLIDYDCLCTDVGDKHCEFDINECLSQPCONGATCSQYVNSYTCPCPLGFSGI 1017
Qy 451 NCS-----SRGCKNDVAVCSPPDG---SCNCKAGMHGVDSIR----- 485
Db 1018 NCQTNDDECTESSCLNGGSC--TDGTINGNCSLQAGYKANKRDSNPLCNGATC 1075
Qy 486 -----CPSGTWGFECNL-----TCQCLNGACNTL--DGTCTCAPGMRGKCE- 526
Db 1076 HEONNRYTCHCPSGFTGKCKSEYVDMCGSPCENAGTCSQMKHOPSCSKSADWTGICDY 1135
Qy 527 --LPCODGT--YGLNCAERCD---CSHADGCHPTTGHCRCLPGMSGVHC----- 568
Db 1136 QRTSQDADADRGLSLRQLCNNGTCKDYGNSHV-----CYCSGAYGASYCQKEIDECQSQP 1191

```


OY 569 -----DSVCAEGRMGPNCSL-----PC 585
 Db 1192 CONGTCDLIGATGECQCRQGFQGNCELINDCAPNFC 1230

RESULT 14 709070

probable tenascin X - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jan-2000
 C:Accession: 709070
 R:Rowen, L.; Mahatras, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Sc submitted to the EMBL Data Library, October 1997
 A:Description: Sequence of the mouse major histocompatibility locus class III region.
 A:Reference number: 216543
 A:Accession: 709070
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4006 <ROM>
 A:Cross-references: EMBL:AF030001; NID:g2564945; PID:g2564958
 C:Genetics:
 A:Gene: TNX
 A:Map position: 17
 A:Initrions: 124/1; 735/1; 826/1; 914/1; 1037/1; 1135/1; 1232/1; 1329/1; 1440/1; 15 019/1; 3115/1; 3208/1; 3302/1; 3405/1; 3517/1; 3558/1; 3606/1; 3646/1; 3694/1; 3737/5; 3 C:Superfamily: tenascin-X; EGF homology; fibrinogen beta/gamma homology; fibronectin ty F:432-448/Domain: EGF homology <EGF>
 F:826-906/Domain: fibrinogen type III repeat homology <3FR>
 F:3789-3997/Domain: fibrinogen beta/gamma homology <Fag>

Query Match 18.88; Score 676.5; DB 2; Length 4006;
 Best Local Similarity 29.18; Pred. No. 3.5e-30;

Matches 190; Conservative 50; Mismatches 178; Indels 235; Gaps 38;

OY 82 HGEKTYRRKSCQCGF-----YESGEMCYPHCAD--KCYHGRCIAPNT 123
 Db 135 HGVEDLSRCACSCPEGMGPTCDPTDKTPTSSPSPKSCDECDNDCRCRGKCY----- 190
 OY 124 CQCEPFGNGTNC--SACDGDHMGHCTSRCCCKNGALCNPTTGACHCAAGFRGMRCEDR 182
 Db 191 --CPFGTSGPSCWSPSGPD-----CQGRGRC-----YKGVCAVCAAGSG--PDC 231
 OY 183 EGTGYGNDCHQRCQCGATCDHYTGRCRCPGYTGAF--EDLPPGKHGQCQRCCQ 241
 Db 232 SQRCSPRNCNDRGCE-----GRVCYCDPGTSGEDCGVRSCEP-----CSQRGCE 278
 OY 242 NGVCHHTVTECCSPSGMGTVC--QCPREGFRGKNCSECCQCHNGGTCDATGQCHSP 300
 Db 279 N-----GLCVNPGYSGEDCGVRNCPRG-----CSQRGCE-----GRVCYCD 317
 OY 301 GYTGRCODECPVGTGYLCAETC--OCVNGKCTHYSGACLCAGFAGERCEARLCEP 358
 Db 318 GYSEEDS-----MRTCPMDCGDGGRG--VDGRVCVCPWPGYSGEDCSYRTPR 363
 OY 359 LYG1-KC-DKRPCHE-----NTHSC-----HPMSECCAKRGMSLYCNE----- 398
 Db 364 CRGRGRDEDECTDAGTSSDDCGVRSCEPDCNDRGCHCEBRCRCMFGYTGADCTACR 423
 OY 399 -----TCSPGFYGAC--QOIC--SCQNGADDSYVGTCTCAAFGKIDCT 441
 Db 424 RDCRGRCEDGVCAHGYSGEDCGVRSCEPDCRGNKNCES--GRVCMPGYTGRCCT 481
 OY 442 -----PPLGTGTGNCSS--RC--GCKADAVCSPYDSCTCTKAGHGV 480
 Db 482 RACPGDCRGGRGVDCVRCVCPNPGTGEDCGSRRCPGDCRGHGE--NGVCVCAVAGSGD 539
 OY 481 DCSIR--CPSGTWGFGCLTQCLNG-----GACNTLDGTC 514
 Db 540 DCSIRCPSPSCRGK-----QCLNGLCEDDEGYSGEDCGTRCPDCSOGHGO--DGLC 592
 OY 515 TCAPGMGKCEKEL-----PCDDGTGTLN-----CAER--CDCSHADGCHPT 553

Db 593 MCHAGTAGEDCSIRTCPADCRRRRCEDGRVCNCGYTGPAACATTCPADCRGRC--V 650
 OY 554 TGHCRCLPQWMSGVHC-----DSVCAEGRMGPNCSL 583
 Db 651 QGVCMCYGYSGEDCGQERPPASACPGCGPRELCRAQCVCVEGFRPDCAI 703

RESULT 15 S18188

notch protein homolog - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 02-Aug-2002
 C:Accession: S18188
 R:Weinmaster, G.; Roberts, V.J.; Lemke, G.
 A:Title: A homolog of Drosophila Notch expressed during mammalian development.
 A:Reference number: S18188; MUID:92111383; PMID:1764995
 A:Accession: S18188
 A:Molecule type: mRNA
 A:Residues: 1-2531 <WEI>
 A:Cross-references: EMBL:X57405; NID:g57634; PID:g57635
 C:Superfamily: notch protein; ankyrin repeat homology; EGF homology
 F:987-1018/Domain: EGF homology <EGF1>
 F:1025-1056/Domain: EGF homology <EGF>
 F:1233-1264/Domain: EGF homology <EGF2>
 F:1917-1949/Domain: ankyrin repeat homology <AN1>
 F:1950-1982/Domain: ankyrin repeat homology <AN2>
 F:1984-2016/Domain: ankyrin repeat homology <AN3>
 F:2017-2049/Domain: ankyrin repeat homology <AN4>
 F:2050-2082/Domain: ankyrin repeat homology <AN5>

Query Match 18.78; Score 675; DB 2; Length 2531;
 Best Local Similarity 25.78; Pred. No. 3.2e-30;

Matches 208; Conservative 70; Mismatches 206; Indels 324; Gaps 48;

OY 4 SLNCSL-----FICLLCHWIGTASPLNEDPNCVSHWESYTVQESY 48
 Db 603 NINECHSOPCRHNGGTQDRDNYVLCILKGTTPNCETIND-----CA----- 646
 OY 49 PHPDQIYVTSCTDILNMFKCTRRHVSRYTAYRHGKTYRRKSQCGFYESGPMC--- 105
 Db 647 SNPCDS--GTCCLKIDGIECA-----CEGTY--TSMCAVN 678
 OY 106 VPBCA-----DKCVHGRG--IAPNT 123
 Db 679 IDECAGSPCHNGTCEDEGIAGFTCRCEGYHDPTCLSEVNECNPNCTHGCACRGLNGYK 738
 OY 124 CQCEPFGNGTNC--SSACDGDHMGHCTSRCCCKNGALCNPTTG--ACHCAAGFRGMR 178
 Db 739 CDCAPGWSGTICDINNNECESN-----PCVNGGTCKDMTSGYVCTCRGFGSPNC 788
 OY 179 EDRCEQTYGNDCHQRCQCGATC--DHYTG--ECRCPPGYTGAFCEBDCPPRGKHPCEQ 236
 Db 789 Q-----TINIEGASN--PCLNQGTCTIDVAGYKNCNCLPTTGACVEYVLA--C-A 834
 OY 237 RCPQNGVC--HVTGECSCPSGMGTVC-----GQPCREG-- 272
 Db 835 TSPCKNSGVKESDEYSEFCVCPMGQGTCEIDINECVKSPCRHGAASCONTNGSYRCL 894
 OY 273 -----FGKNS--QEQ-----CHNGTCDAATG--CHCSPGTGECODE----- 310
 Db 895 CQAGYTRNCESDIDCPNPNCHNGSGCTDGVNAAFCDCLPFGGACCEBDINECATNPC 954
 OY 311 -----CPVGTGYLCAET-----CCQVNGKCYHNSG-----ACLCGAG 344
 Db 955 QNAGNCTDVCDSYTCPTFPNGIHCENNPTDCTESSCFNGGTC--VDGINSFCLCPPG 1012
 OY 345 FAGECEARLCPGLYGI--KCDKRPCHELENTSHCHPMG--ECACKPGMSGLYCN-- 398
 Db 1013 FTGSGYQ-----YDVNCDISR--PCLHGGT--CQDSYGTCTGCTGGLMCCQLVLR 1061
 OY 399 --TCSPGFYGACQOI-----CSCQN--GADDSDYTGKCTCAPRGFGIDCSITPCLGTY 448

```

Db 1062 WCDAPCKNGKCMQNTMTQYHCECRSGMTGFNCVLSVSCVFAAQRGIDVTLLCGHGL 1121
QY 449 GIN-----CSSRCG-----CKNDVAVSPVDG--SCTKAGMHGVDCS 483
Db 1122 CYDEEDRHHYCHQAGYGSYCEDEVEDCSPNPNCONGATCFDYLGGFSCKCVAGYHGSNCS 1181
QY 484 -----IRPSTGTMGFGCNLT--C-----QCLNG 504
Db 1182 BEINECLSPQCONGGTCTIDLTNTYKSCSPRGTOGVHCEINVDCHPPLDPASRSPKCFNN 1241
QY 505 GACNTLDG--TCTCAPGMGEKCE-----LPCOD-GTYGLNCAERDCSHADGCHPT 553
Db 1242 GTCVDQYGGYTCCTCPGFVGBERCEGDVNECLSNPCDPRGTQ--NCVQRYN----- 1289
QY 554 TGHCRCLPGWSGVHCDVSCAEGRMGPNC 581
Db 1290 DFHCECRAGHTGRRCESV-INGCRGKPC 1316

```

Search completed: May 9, 2003, 11:58:44
 Job time : 34.9757 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 9, 2003, 10:55:56 ; Search time 9.50637 Seconds

(Without alignments)
2556.719 Million cell updates/sec

File: US-10-092-390-4
Database: 3601
Sequence: 1 MWISLNSCLSFICLCHWL.....HCDSCVCAEGRMGPCSLPCY 586

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_40

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	843.5	23.4	1 SREC_HUMAN	Q14162 homo sapien
2	719	20.0	1 NOTC_XENLA	P21783 xenopus lae
3	717	19.9	1 NTCL_BRARE	P46550 brachydanio
4	697	19.4	1 NTCL_HUMAN	Q04721 homo sapien
5	693	19.2	1 NTCL_MOUSE	O01705 mus musculu
6	685.5	19.0	1 NTCL_MOUSE	O61982 mus musculu
7	685.5	19.0	1 NTCL_MOUSE	O61982 mus musculu
8	682.5	19.0	1 NTCL_MOUSE	O61982 mus musculu
9	682.5	19.0	1 NTCL_MOUSE	O61982 mus musculu
10	677	18.8	1 TENX_HUMAN	Q91172 ratu
11	677	18.8	1 NOTC_MOUSE	P22105 homo sapien
12	675.5	18.8	1 NOTC_MOUSE	P22105 homo sapien
13	675.5	18.8	1 NOTC_MOUSE	P22105 homo sapien
14	667.5	18.5	1 JAG2_HUMAN	O07008 ratu
15	666.5	18.5	1 JAG2_HUMAN	O07008 ratu
16	664.5	18.5	1 FBPI_STRPU	Q04721 homo sapien
17	662	18.4	1 NTCL_MOUSE	O01705 mus musculu
18	658	18.3	1 NTCL_MOUSE	O01705 mus musculu
19	646	17.9	1 TENA_CHICK	P16593 mus musculu
20	644.5	17.9	1 TENA_CHICK	P16593 mus musculu
21	638	17.7	1 LMA5_HUMAN	P10033 gallu
22	633	17.6	1 JAG2_HUMAN	O07008 ratu
23	631.5	17.5	1 JAG2_HUMAN	O07008 ratu
24	631	17.5	1 JAG2_HUMAN	O07008 ratu
25	629	17.5	1 JAG2_HUMAN	O07008 ratu
26	619.5	17.2	1 JAG2_HUMAN	O07008 ratu
27	616	17.1	1 JAG2_HUMAN	O07008 ratu
28	615	17.1	1 JAG2_HUMAN	O07008 ratu
29	614	17.1	1 JAG2_HUMAN	O07008 ratu
30	611	17.0	1 JAG2_HUMAN	O07008 ratu
31	593.5	16.5	1 LML2_CAEBL	Q21313 caenorhabdi
32	587	16.3	1 LML2_CAEBL	Q21313 caenorhabdi
33	586	16.3	1 LMA5_MOUSE	O61001 mus musculu

34	577.5	16.0	1 LMB2_RAT	P15800 ratu
35	576.5	16.0	1 LMA_DROME	O00174 drosophila
36	567	15.7	1 LMB2_MOUSE	P55268 homo sapien
37	564.5	15.7	1 LMB2_MOUSE	O61292 mus musculu
38	561	15.6	1 LMA2_MOUSE	O60675 mus musculu
39	560.5	15.6	1 LMA2_MOUSE	P14585 caenorhabdi
40	559	15.5	1 LMA2_HUMAN	P24043 homo sapien
41	556	15.4	1 DL_DROME	P10041 drosophila
42	546.5	15.2	1 LMA1_MOUSE	P19137 mus musculu
43	536	14.9	1 LMA1_MOUSE	P19137 mus musculu
44	533	14.8	1 LMA1_MOUSE	P19137 mus musculu
45	517.5	14.4	1 LMB1_MOUSE	P02469 mus musculu

ALIGNMENTS

RESULT 1	SREC_HUMAN	STANDARD	PRT	830 AA.
AC	Q14162: 043701;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Endothelial cells scavenger receptor precursor (Acetyl LDL receptor).			
GN	SREC OR KIAA0149.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX	NCBI_Taxid:9606;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Umbilical vein endothelial cells;			
KX	MEDLINE=9805897; PubMed=9395444;			
RA	Adachi H., Tsujimoto M., Arai H., Inoue K.;			
RT	"Expression cloning of a novel scavenger receptor from human			
RL	endothelial cells."			
RL	J. Biol. Chem. 272:31217-31220(1997).			
LN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Bone marrow;			
KX	MEDLINE=96127530; PubMed=8590280;			
RA	Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;			
RT	"Prediction of the coding sequences of unidentified human genes. IV.			
RT	The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by			
RT	analysis of cDNA clones from human cell line KG-1."			
RL	DNA Res. 2:167-174(1995).			
CC	-1- FUNCTION: Mediates the binding and degradation of acetylated low			
CC	density lipoprotein (Ac-LDL).			
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).			
CC	-1- TISSUE SPECIFICITY: Endothelial cells.			
CC	-1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: D63864; BAA24070.1; -			
DR	EMBL: D63483; BAA09770.1; -			
DR	HSSP: P01180; 2B2.			
DR	InterPro: IPR000561; EGF-like.			
DR	PROSITE: PS00022; EGF_1; 6.			
KW	Receptor; Glycoprotein; Signal; Transmembrane; EGF-like domain;			
KW	Repeat.			
FT	SIGNAL	1	19	POTENTIAL.
FT	CHAIN	20	830	ENDOTHELIAL CELLS SCAVENGER RECEPTOR.
FT	DOMAIN	20	421	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	422	442	POTENTIAL.

```

FT DOMAIN 443 830 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 53 87 EGF-LIKE 1.
FT DOMAIN 95 130 EGF-LIKE 2.
FT DOMAIN 155 191 EGF-LIKE 3.
FT DOMAIN 215 249 EGF-LIKE 4.
FT DOMAIN 302 339 EGF-LIKE 5.
FT DOMAIN 351 382 EGF-LIKE 6.
FT DOMAIN 476 620 PRO/SER-RICH.
FT DOMAIN 622 798 GLY-RICH.
FT DOMAIN 431 438 POLY-LEU.
FT DISULFID 57 69 POTENTIAL.
FT DISULFID 63 75 POTENTIAL.
FT DISULFID 77 86 POTENTIAL.
FT DISULFID 99 111 POTENTIAL.
FT DISULFID 105 118 POTENTIAL.
FT DISULFID 120 129 POTENTIAL.
FT DISULFID 159 172 POTENTIAL.
FT DISULFID 165 179 POTENTIAL.
FT DISULFID 181 190 POTENTIAL.
FT DISULFID 219 230 POTENTIAL.
FT DISULFID 225 237 POTENTIAL.
FT DISULFID 239 248 POTENTIAL.
FT DISULFID 306 319 POTENTIAL.
FT DISULFID 313 326 POTENTIAL.
FT DISULFID 329 338 POTENTIAL.
FT DISULFID 335 363 POTENTIAL.
FT DISULFID 358 370 POTENTIAL.
FT DISULFID 372 381 POTENTIAL.
FT CAROHND 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROHND 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 662 662 R -> W (IN REF. 2).
SQ SEQUENCE 830 AA; 87430 MM; F560D9ELA64D779 CRC64;

```

```

Query Match 23.4%; Score 843.5; DB 1; Length 830;
Best Local Similarity 36.2%; Pred. No. 7,6e-47;
Matches 158; Conservative 51; Mismatches 164; Indels 63; Gaps 15;

```

```

QY 93 OCCPGFSEGEHC-VPHC--ADKCYHGR-CIAPNTQCCEPFGMGNTCCSACDGDHMGPHC 148
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 40 OCCACMROROKDOCTPIICGPRACQKDEVCVRPGLCRKRFEGAHCSRCRPGQYWGPDG 99
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 TNRCCCKNALNPITGACHCAAGFRGWRCEDECEGTGNDCHQCCOONATCDHYTG 208
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 RESCPHGHGCEPATGACQCADRWGARGCEPPCAGPHGR-----CDPATG 146
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 209 ECRCPFGYTGAFCEDLCPGKKGPOCEORPCONGGVCHHTGECSCPSGMMGTVCQGPC 268
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 147 VCHCEGKMSSSTCRPRCCQNTAAARCEQ-----ATGACVCKRGW----- 185
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 PGRGRGKNCQSCQCHNGSTDAATGQCHCSFGYTGERCQDECPVGTGYVLAETGQCVN 328
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 186 ----WGRRCISFRGNCH-GSPCEQDSRCACRPGMWGPECQO-----CECYR 227
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 329 GSKCYHVSACLCCEAGFNGERCEARLCPGLGKICDKRCPHLENTHSCHPMSGEC-AC 387
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 228 -GRCSAASECCTCPGFRARCELP-CRPSHGVCQAHSCG-RCKINPESCSPTGSCESC 284
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 KPGMSGLVCNETSPGFGYEGACQOIC-SCQNGADCDVYTK-TCAPGFKGIDSTPCPL 445
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 285 ECGHNGTQCCQCLPCTGTRESCEDQCPHRHGEACEPDTGKHORCDPGLGRCCEPCT 344
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 GTYGINCSSRCCKNDAYCSPYDGSCTCAAGHGVDCSIRCPSGTWGFGCNTLTCQCLNG 505
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 345 GFGEDECGSTCPTVOGSDVTGDCVCSAGYWGSPSCNMSCPAGFHGNNCSVPCCEPE-G 403
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 506 ACNTLDGTCTCAPGR 521
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 404 LCHPVSGSCOPSGSGR 419
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 2
NOTC_XENLA

```

ID NOTC_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog precursor (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae;
OC Xenopodinae; Xenopus.
OX NCBI_Taxid=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; PubMed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Xotch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M3874; AB02039.1; -
CC PIR; A35844; A35844.
CC HSP; P00740; 1EDM.
CC InterPro: IPR002110; ANK.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-Like.
CC InterPro: IPR00742; EGF-2.
CC InterPro: IPR001811; EGF_CA.
CC InterPro: IPR001438; EGF_IL.
CC InterPro: IPR000800; Notch.
CC Pfam; PF00008; EGF; 36.
CC Pfam; PF00023; ank; 6.
CC PRINTS; PR00010; EGFBL0D.
CC PRINTS; PR01452; NOTCH.
CC SMART; SM00248; ANK; 5.
CC SMART; SM00179; EGF_CA; 23.
CC SMART; SM00001; EGF-Like; 11.
CC SMART; SM00004; NL; 2.
CC PROSITE; PS50088; ANK_REPEAT; 4.
CC PROSITE; PS50297; ANK_REPEAT_REGION; 1.
CC PROSITE; PS00010; ASX_HYDROXYL; 23.
CC PROSITE; PS00022; EGF_1; 34.
CC PROSITE; PS01186; EGF_2; 29.
CC PROSITE; PS01187; EGF_CA; 21.
CC Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 2524
FT DOMAIN 20 1728
FT TRANSMEM 1729 1750
FT DOMAIN 1751 2524
FT DOMAIN 20 57
FT DOMAIN 58 99
FT DOMAIN 102 140
FT DOMAIN 141 177
FT EGF-LIKE 4.

```

```
FT DOMAIN 179 215 EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 217 254 EGF-LIKE 6.
FT DOMAIN 256 292 EGF-LIKE 7, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 284 332 EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 334 370 EGF-LIKE 9, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 371 409 EGF-LIKE 10.
FT DOMAIN 411 449 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 451 487 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 489 525 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 527 563 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 565 600 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 602 638 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 640 675 EGF-LIKE 17.
FT DOMAIN 677 713 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 715 750 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 752 788 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 790 826 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 828 866 EGF-LIKE 22.
FT DOMAIN 868 904 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 906 942 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 944 980 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 982 1018 EGF-LIKE 26.
FT DOMAIN 1020 1056 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1058 1094 EGF-LIKE 28.
FT DOMAIN 1096 1142 EGF-LIKE 29.
FT DOMAIN 1144 1180 EGF-LIKE 30.
FT DOMAIN 1182 1218 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1220 1264 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1266 1304 EGF-LIKE 33.
FT DOMAIN 1306 1346 EGF-LIKE 34.
FT DOMAIN 1347 1383 EGF-LIKE 35.
FT DOMAIN 1386 1424 EGF-LIKE 36.
FT REPEAT 1441 1478 LIN/NORCH 1.
FT REPEAT 1479 1520 LIN/NORCH 2.
FT REPEAT 1521 1560 LIN/NORCH 3.
FT REPEAT 1560 1594 ANK 1.
FT REPEAT 1594 1626 ANK 2.
FT REPEAT 1626 1658 ANK 3.
FT REPEAT 1658 1690 ANK 4.
FT REPEAT 1690 1722 ANK 5.
FT REPEAT 1722 1754 ANK 6.
FT DISULFID 1754 1786 BY SIMILARITY.
FT DISULFID 1786 1818 BY SIMILARITY.
FT DISULFID 1818 1850 BY SIMILARITY.
FT DISULFID 1850 1882 BY SIMILARITY.
FT DISULFID 1882 1914 BY SIMILARITY.
FT DISULFID 1914 1946 BY SIMILARITY.
FT DISULFID 1946 1978 BY SIMILARITY.
FT DISULFID 1978 2010 BY SIMILARITY.
FT DISULFID 2010 2042 BY SIMILARITY.
FT DISULFID 2042 2074 BY SIMILARITY.
FT DISULFID 2074 2106 BY SIMILARITY.
FT DISULFID 2106 2138 BY SIMILARITY.
FT DISULFID 2138 2170 BY SIMILARITY.
FT DISULFID 2170 2202 BY SIMILARITY.
FT DISULFID 2202 2234 BY SIMILARITY.
FT DISULFID 2234 2266 BY SIMILARITY.
FT DISULFID 2266 2298 BY SIMILARITY.
FT DISULFID 2298 2330 BY SIMILARITY.
FT DISULFID 2330 2362 BY SIMILARITY.
FT DISULFID 2362 2394 BY SIMILARITY.
FT DISULFID 2394 2426 BY SIMILARITY.
FT DISULFID 2426 2458 BY SIMILARITY.
FT DISULFID 2458 2490 BY SIMILARITY.
FT DISULFID 2490 2522 BY SIMILARITY.
FT DISULFID 2522 2554 BY SIMILARITY.
FT DISULFID 2554 2586 BY SIMILARITY.
FT DISULFID 2586 2618 BY SIMILARITY.
FT DISULFID 2618 2650 BY SIMILARITY.
FT DISULFID 2650 2682 BY SIMILARITY.
FT DISULFID 2682 2714 BY SIMILARITY.
FT DISULFID 2714 2746 BY SIMILARITY.
FT DISULFID 2746 2778 BY SIMILARITY.
FT DISULFID 2778 2810 BY SIMILARITY.
FT DISULFID 2810 2842 BY SIMILARITY.
FT DISULFID 2842 2874 BY SIMILARITY.
FT DISULFID 2874 2906 BY SIMILARITY.
FT DISULFID 2906 2938 BY SIMILARITY.
FT DISULFID 2938 2970 BY SIMILARITY.
FT DISULFID 2970 3002 BY SIMILARITY.
FT DISULFID 3002 3034 BY SIMILARITY.
FT DISULFID 3034 3066 BY SIMILARITY.
FT DISULFID 3066 3098 BY SIMILARITY.
FT DISULFID 3098 3130 BY SIMILARITY.
FT DISULFID 3130 3162 BY SIMILARITY.
FT DISULFID 3162 3194 BY SIMILARITY.
FT DISULFID 3194 3226 BY SIMILARITY.
FT DISULFID 3226 3258 BY SIMILARITY.
FT DISULFID 3258 3290 BY SIMILARITY.
FT DISULFID 3290 3322 BY SIMILARITY.
FT DISULFID 3322 3354 BY SIMILARITY.
FT DISULFID 3354 3386 BY SIMILARITY.
FT DISULFID 3386 3418 BY SIMILARITY.
FT DISULFID 3418 3450 BY SIMILARITY.
FT DISULFID 3450 3482 BY SIMILARITY.
FT DISULFID 3482 3514 BY SIMILARITY.
FT DISULFID 3514 3546 BY SIMILARITY.
FT DISULFID 3546 3578 BY SIMILARITY.
FT DISULFID 3578 3610 BY SIMILARITY.
FT DISULFID 3610 3642 BY SIMILARITY.
FT DISULFID 3642 3674 BY SIMILARITY.
FT DISULFID 3674 3706 BY SIMILARITY.
FT DISULFID 3706 3738 BY SIMILARITY.
FT DISULFID 3738 3770 BY SIMILARITY.
FT DISULFID 3770 3802 BY SIMILARITY.
FT DISULFID 3802 3834 BY SIMILARITY.
FT DISULFID 3834 3866 BY SIMILARITY.
FT DISULFID 3866 3898 BY SIMILARITY.
FT DISULFID 3898 3930 BY SIMILARITY.
FT DISULFID 3930 3962 BY SIMILARITY.
FT DISULFID 3962 3994 BY SIMILARITY.
FT DISULFID 3994 4026 BY SIMILARITY.
FT DISULFID 4026 4058 BY SIMILARITY.
FT DISULFID 4058 4090 BY SIMILARITY.
FT DISULFID 4090 4122 BY SIMILARITY.
```

```
FT DISULFID 439 448 BY SIMILARITY.
FT DISULFID 448 455 BY SIMILARITY.
FT DISULFID 455 466 BY SIMILARITY.
FT DISULFID 466 475 BY SIMILARITY.
FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 486 493 BY SIMILARITY.
FT DISULFID 493 504 BY SIMILARITY.
FT DISULFID 504 513 BY SIMILARITY.
FT DISULFID 513 524 BY SIMILARITY.
FT DISULFID 524 531 BY SIMILARITY.
FT DISULFID 531 542 BY SIMILARITY.
FT DISULFID 542 551 BY SIMILARITY.
FT DISULFID 551 559 BY SIMILARITY.
FT DISULFID 559 562 BY SIMILARITY.
FT DISULFID 562 572 BY SIMILARITY.
FT DISULFID 572 588 BY SIMILARITY.
FT DISULFID 588 599 BY SIMILARITY.
FT DISULFID 599 606 BY SIMILARITY.
FT DISULFID 606 617 BY SIMILARITY.
FT DISULFID 617 626 BY SIMILARITY.
FT DISULFID 626 637 BY SIMILARITY.
FT DISULFID 637 644 BY SIMILARITY.
FT DISULFID 644 654 BY SIMILARITY.
FT DISULFID 654 663 BY SIMILARITY.
FT DISULFID 663 674 BY SIMILARITY.
FT DISULFID 674 681 BY SIMILARITY.
FT DISULFID 681 692 BY SIMILARITY.
FT DISULFID 692 701 BY SIMILARITY.
FT DISULFID 701 712 BY SIMILARITY.
FT DISULFID 712 729 BY SIMILARITY.
FT DISULFID 729 738 BY SIMILARITY.
FT DISULFID 738 749 BY SIMILARITY.
FT DISULFID 749 756 BY SIMILARITY.
FT DISULFID 756 767 BY SIMILARITY.
FT DISULFID 767 776 BY SIMILARITY.
FT DISULFID 776 787 BY SIMILARITY.
FT DISULFID 787 805 BY SIMILARITY.
FT DISULFID 805 814 BY SIMILARITY.
FT DISULFID 814 825 BY SIMILARITY.
FT DISULFID 825 832 BY SIMILARITY.
FT DISULFID 832 843 BY SIMILARITY.
FT DISULFID 843 854 BY SIMILARITY.
FT DISULFID 854 865 BY SIMILARITY.
FT DISULFID 865 872 BY SIMILARITY.
FT DISULFID 872 883 BY SIMILARITY.
FT DISULFID 883 892 BY SIMILARITY.
FT DISULFID 892 903 BY SIMILARITY.
FT DISULFID 903 921 BY SIMILARITY.
FT DISULFID 921 930 BY SIMILARITY.
FT DISULFID 930 941 BY SIMILARITY.
FT DISULFID 941 952 BY SIMILARITY.
FT DISULFID 952 966 BY SIMILARITY.
FT DISULFID 966 977 BY SIMILARITY.
FT DISULFID 977 1006 BY SIMILARITY.
FT DISULFID 1006 1017 BY SIMILARITY.
FT DISULFID 1017 1035 BY SIMILARITY.
FT DISULFID 1035 1044 BY SIMILARITY.
FT DISULFID 1044 1055 BY SIMILARITY.
FT DISULFID 1055 1062 BY SIMILARITY.
FT DISULFID 1062 1073 BY SIMILARITY.
FT DISULFID 1073 1082 BY SIMILARITY.
FT DISULFID 1082 1093 BY SIMILARITY.
FT DISULFID 1093 1121 BY SIMILARITY.
FT DISULFID 1121 1130 BY SIMILARITY.
FT DISULFID 1130 1141 BY SIMILARITY.
FT DISULFID 1141 1159 BY SIMILARITY.
FT DISULFID 1159 1168 BY SIMILARITY.
```

Query Match 20.0%: Score 719; DB 1; Length 2524;

Best Local Similarity 25.8%; Pred No. 1.3e-38; Mismatches 222; Indels 355; Gaps 50;

Matches 222; Conservative 60;

```
QY 5 INSCISFICL-----LICHMIGTASPLNED-----PNVCSHM-----ESYS 41
DB 603 INECLSKRCLNGGQCTDRENGVYICPKGTGVNCEKIDDCASMLCNGKIDKIDGVE 662
QY 42 VTVOESYRPHFDQIYTT-----SCDILAMFKCTRRVSRFAYRGEKTYRR 90
DB 663 CYCEPGYTGKLCININEDCSNPCRNGCYCKXQINGFCV----- 702
QY 91 KQCCPGYRSESGMC---VPHC-ADKCVHRC---IAPWCCCEPFGMGFTNC---SSACD 140
DB 703 ---CPDGYHD-HMCLSEVNECNSNPCILGACHDGVNGYKCCCEAGSGSNGDINNNECE 757
QY 141 GDHMGPHCTSRCCAKNGALCNPIITGA--CHCAAGFRGRCEDRCBOGTYYGNDCHQRCOCO 198
```



```

FT DOMAIN 789 825 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 865 865 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 867 903 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 905 941 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 943 979 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 981 1017 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1019 1055 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1057 1093 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1095 1141 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1143 1179 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1181 1217 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1219 1263 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1265 1303 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1305 1344 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1346 1382 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1385 1423 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 1446 1486 LIN/NOTCH 1.
FT REPEAT 1487 1520 LIN/NOTCH 2.
FT REPEAT 1521 1561 LIN/NOTCH 3.
FT REPEAT 1562 1602 ANK 1.
FT REPEAT 1603 1643 ANK 2.
FT REPEAT 1644 1684 ANK 3.
FT REPEAT 1685 1725 ANK 4.
FT REPEAT 1726 1766 ANK 5.
FT REPEAT 1767 1807 ANK 6.
FT REPEAT 1808 1848 ANK 7.
FT REPEAT 1849 1889 ANK 8.
FT REPEAT 1890 1930 ANK 9.
FT REPEAT 1931 1971 ANK 10.
FT REPEAT 1972 2012 ANK 11.
FT REPEAT 2013 2053 ANK 12.
FT REPEAT 2054 2094 ANK 13.
FT REPEAT 2095 2135 ANK 14.
FT REPEAT 2136 2176 ANK 15.
FT REPEAT 2177 2217 ANK 16.
FT REPEAT 2218 2258 ANK 17.
FT REPEAT 2259 2299 ANK 18.
FT REPEAT 2300 2340 ANK 19.
FT REPEAT 2341 2381 ANK 20.
FT REPEAT 2382 2422 ANK 21.
FT REPEAT 2423 2463 ANK 22.
FT REPEAT 2464 2504 ANK 23.
FT REPEAT 2505 2545 ANK 24.
FT REPEAT 2546 2586 ANK 25.
FT REPEAT 2587 2627 ANK 26.
FT REPEAT 2628 2668 ANK 27.
FT REPEAT 2669 2709 ANK 28.
FT REPEAT 2710 2750 ANK 29.
FT REPEAT 2751 2791 ANK 30.
FT REPEAT 2792 2832 ANK 31.
FT REPEAT 2833 2873 ANK 32.
FT REPEAT 2874 2914 ANK 33.
FT REPEAT 2915 2955 ANK 34.
FT REPEAT 2956 2996 ANK 35.
FT REPEAT 2997 3037 ANK 36.
FT REPEAT 3038 3078 ANK 37.
FT REPEAT 3079 3119 ANK 38.
FT REPEAT 3120 3160 ANK 39.
FT REPEAT 3161 3201 ANK 40.
FT REPEAT 3202 3242 ANK 41.
FT REPEAT 3243 3283 ANK 42.
FT REPEAT 3284 3324 ANK 43.
FT REPEAT 3325 3365 ANK 44.
FT REPEAT 3366 3406 ANK 45.
FT REPEAT 3407 3447 ANK 46.
FT REPEAT 3448 3488 ANK 47.
FT REPEAT 3489 3529 ANK 48.
FT REPEAT 3530 3570 ANK 49.
FT REPEAT 3571 3611 ANK 50.
FT REPEAT 3612 3652 ANK 51.
FT REPEAT 3653 3693 ANK 52.
FT REPEAT 3694 3734 ANK 53.
FT REPEAT 3735 3775 ANK 54.
FT REPEAT 3776 3816 ANK 55.
FT REPEAT 3817 3857 ANK 56.
FT REPEAT 3858 3898 ANK 57.
FT REPEAT 3899 3939 ANK 58.
FT REPEAT 3940 3980 ANK 59.
FT REPEAT 3981 4021 ANK 60.
FT REPEAT 4022 4062 ANK 61.
FT REPEAT 4063 4103 ANK 62.
FT REPEAT 4104 4144 ANK 63.
FT REPEAT 4145 4185 ANK 64.
FT REPEAT 4186 4226 ANK 65.
FT REPEAT 4227 4267 ANK 66.
FT REPEAT 4268 4308 ANK 67.
FT REPEAT 4309 4349 ANK 68.
FT REPEAT 4350 4390 ANK 69.
FT REPEAT 4391 4431 ANK 70.
FT REPEAT 4432 4472 ANK 71.
FT REPEAT 4473 4513 ANK 72.
FT REPEAT 4514 4554 ANK 73.
FT REPEAT 4555 4595 ANK 74.
FT REPEAT 4596 4636 ANK 75.
FT REPEAT 4637 4677 ANK 76.
FT REPEAT 4678 4718 ANK 77.
FT REPEAT 4719 4759 ANK 78.
FT REPEAT 4760 4800 ANK 79.
FT REPEAT 4801 4841 ANK 80.
FT REPEAT 4842 4882 ANK 81.
FT REPEAT 4883 4923 ANK 82.
FT REPEAT 4924 4964 ANK 83.
FT REPEAT 4965 5005 ANK 84.
FT REPEAT 5006 5046 ANK 85.
FT REPEAT 5047 5087 ANK 86.
FT REPEAT 5088 5128 ANK 87.
FT REPEAT 5129 5169 ANK 88.
FT REPEAT 5170 5210 ANK 89.
FT REPEAT 5211 5251 ANK 90.
FT REPEAT 5252 5292 ANK 91.
FT REPEAT 5293 5333 ANK 92.
FT REPEAT 5334 5374 ANK 93.
FT REPEAT 5375 5415 ANK 94.
FT REPEAT 5416 5456 ANK 95.
FT REPEAT 5457 5497 ANK 96.
FT REPEAT 5498 5538 ANK 97.
FT REPEAT 5539 5579 ANK 98.
FT REPEAT 5580 5620 ANK 99.
FT REPEAT 5621 5661 ANK 100.
FT REPEAT 5662 5702 ANK 101.
FT REPEAT 5703 5743 ANK 102.
FT REPEAT 5744 5784 ANK 103.
FT REPEAT 5785 5825 ANK 104.
FT REPEAT 5826 5866 ANK 105.
FT REPEAT 5867 5907 ANK 106.
FT REPEAT 5908 5948 ANK 107.
FT REPEAT 5949 5989 ANK 108.
FT REPEAT 5990 6030 ANK 109.
FT REPEAT 6031 6071 ANK 110.
FT REPEAT 6072 6112 ANK 111.
FT REPEAT 6113 6153 ANK 112.
FT REPEAT 6154 6194 ANK 113.
FT REPEAT 6195 6235 ANK 114.
FT REPEAT 6236 6276 ANK 115.
FT REPEAT 6277 6317 ANK 116.
FT REPEAT 6318 6358 ANK 117.
FT REPEAT 6359 6399 ANK 118.
FT REPEAT 6400 6440 ANK 119.
FT REPEAT 6441 6481 ANK 120.
FT REPEAT 6482 6522 ANK 121.
FT REPEAT 6523 6563 ANK 122.
FT REPEAT 6564 6604 ANK 123.
FT REPEAT 6605 6645 ANK 124.
FT REPEAT 6646 6686 ANK 125.
FT REPEAT 6687 6726 ANK 126.
FT REPEAT 6727 6767 ANK 127.
FT REPEAT 6768 6808 ANK 128.
FT REPEAT 6809 6849 ANK 129.
FT REPEAT 6850 6890 ANK 130.
FT REPEAT 6891 6931 ANK 131.
FT REPEAT 6932 6972 ANK 132.
FT REPEAT 6973 7013 ANK 133.
FT REPEAT 7014 7054 ANK 134.
FT REPEAT 7055 7095 ANK 135.
FT REPEAT 7096 7136 ANK 136.
FT REPEAT 7137 7177 ANK 137.
FT REPEAT 7178 7218 ANK 138.
FT REPEAT 7219 7259 ANK 139.
FT REPEAT 7260 7300 ANK 140.
FT REPEAT 7301 7341 ANK 141.
FT REPEAT 7342 7382 ANK 142.
FT REPEAT 7383 7423 ANK 143.
FT REPEAT 7424 7464 ANK 144.
FT REPEAT 7465 7505 ANK 145.
FT REPEAT 7506 7546 ANK 146.
FT REPEAT 7547 7587 ANK 147.
FT REPEAT 7588 7628 ANK 148.
FT REPEAT 7629 7669 ANK 149.
FT REPEAT 7670 7710 ANK 150.
FT REPEAT 7711 7751 ANK 151.
FT REPEAT 7752 7792 ANK 152.
FT REPEAT 7793 7833 ANK 153.
FT REPEAT 7834 7874 ANK 154.
FT REPEAT 7875 7915 ANK 155.
FT REPEAT 7916 7956 ANK 156.
FT REPEAT 7957 7997 ANK 157.
FT REPEAT 7998 8038 ANK 158.
FT REPEAT 8039 8079 ANK 159.
FT REPEAT 8080 8120 ANK 160.
FT REPEAT 8121 8161 ANK 161.
FT REPEAT 8162 8202 ANK 162.
FT REPEAT 8203 8243 ANK 163.
FT REPEAT 8244 8284 ANK 164.
FT REPEAT 8285 8325 ANK 165.
FT REPEAT 8326 8366 ANK 166.
FT REPEAT 8367 8407 ANK 167.
FT REPEAT 8408 8448 ANK 168.
FT REPEAT 8449 8489 ANK 169.
FT REPEAT 8490 8530 ANK 170.
FT REPEAT 8531 8571 ANK 171.
FT REPEAT 8572 8612 ANK 172.
FT REPEAT 8613 8653 ANK 173.
FT REPEAT 8654 8694 ANK 174.
FT REPEAT 8695 8735 ANK 175.
FT REPEAT 8736 8776 ANK 176.
FT REPEAT 8777 8817 ANK 177.
FT REPEAT 8818 8858 ANK 178.
FT REPEAT 8859 8899 ANK 179.
FT REPEAT 8900 8940 ANK 180.
FT REPEAT 8941 8981 ANK 181.
FT REPEAT 8982 9022 ANK 182.
FT REPEAT 9023 9063 ANK 183.
FT REPEAT 9064 9104 ANK 184.
FT REPEAT 9105 9145 ANK 185.
FT REPEAT 9146 9186 ANK 186.
FT REPEAT 9187 9227 ANK 187.
FT REPEAT 9228 9268 ANK 188.
FT REPEAT 9269 9309 ANK 189.
FT REPEAT 9310 9350 ANK 190.
FT REPEAT 9351 9391 ANK 191.
FT REPEAT 9392 9432 ANK 192.
FT REPEAT 9433 9473 ANK 193.
FT REPEAT 9474 9514 ANK 194.
FT REPEAT 9515 9555 ANK 195.
FT REPEAT 9556 9596 ANK 196.
FT REPEAT 9597 9637 ANK 197.
FT REPEAT 9638 9678 ANK 198.
FT REPEAT 9679 9719 ANK 199.
FT REPEAT 9720 9760 ANK 200.
FT REPEAT 9761 9801 ANK 201.
FT REPEAT 9802 9842 ANK 202.
FT REPEAT 9843 9883 ANK 203.
FT REPEAT 9884 9924 ANK 204.
FT REPEAT 9925 9965 ANK 205.
FT REPEAT 9966 10006 ANK 206.
FT REPEAT 10007 10047 ANK 207.
FT REPEAT 10048 10088 ANK 208.
FT REPEAT 10089 10129 ANK 209.
FT REPEAT 10130 10170 ANK 210.
FT REPEAT 10171 10211 ANK 211.
FT REPEAT 10212 10252 ANK 212.
FT REPEAT 10253 10293 ANK 213.
FT REPEAT 10294 10334 ANK 214.
FT REPEAT 10335 10375 ANK 215.
FT REPEAT 10376 10416 ANK 216.
FT REPEAT 10417 10457 ANK 217.
FT REPEAT 10458 10498 ANK 218.
FT REPEAT 10499 10539 ANK 219.
FT REPEAT 10540 10580 ANK 220.
FT REPEAT 10581 10621 ANK 221.
FT REPEAT 10622 10662 ANK 222.
FT REPEAT 10663 10703 ANK 223.
FT REPEAT 10704 10744 ANK 224.
FT REPEAT 10745 10785 ANK 225.
FT REPEAT 10786 10825 ANK 226.
FT REPEAT 10826 10866 ANK 227.
FT REPEAT 10867 10906 ANK 228.
FT REPEAT 10907 10947 ANK 229.
FT REPEAT 10948 10988 ANK 230.
FT REPEAT 10989 11029 ANK 231.
FT REPEAT 11030 11070 ANK 232.
FT REPEAT 11071 11111 ANK 233.
FT REPEAT 11112 11152 ANK 234.
FT REPEAT 11153 11193 ANK 235.
FT REPEAT 11194 11234 ANK 236.
FT REPEAT 11235 11274 ANK 237.
FT REPEAT 11275 11315 ANK 238.
FT REPEAT 11316 11356 ANK 239.
FT REPEAT 11357 11396 ANK 240.
FT REPEAT 11397 11437 ANK 241.
FT REPEAT 11438 11478 ANK 242.
FT REPEAT 11479 11518 ANK 243.
FT REPEAT 11519 11559 ANK 244.
FT REPEAT 11560 11600 ANK 245.
FT REPEAT 11601 11641 ANK 246.
FT REPEAT 11642 11682 ANK 247.
FT REPEAT 11683 11722 ANK 248.
FT REPEAT 11723 11763 ANK 249.
FT REPEAT 11764 11803 ANK 250.
FT REPEAT 11804 11844 ANK 251.
FT REPEAT 11845 11885 ANK 252.
FT REPEAT 11886 11925 ANK 253.
FT REPEAT 11926 11966 ANK 254.
FT REPEAT 11967 12006 ANK 255.
FT REPEAT 12007 12047 ANK 256.
FT REPEAT 12048 12088 ANK 257.
FT REPEAT 12089 12129 ANK 258.
FT REPEAT 12130 12170 ANK 259.
FT REPEAT 12171 12211 ANK 260.
FT REPEAT 12212 12252 ANK 261.
FT REPEAT 12253 12293 ANK 262.
FT REPEAT 12294 12334 ANK 263.
FT REPEAT 12335 12375 ANK 264.
FT REPEAT 12376 12416 ANK 265.
FT REPEAT 12417 12457 ANK 266.
FT REPEAT 12458 12498 ANK 267.
FT REPEAT 12499 12539 ANK 268.
FT REPEAT 12540 12580 ANK 269.
FT REPEAT 12581 12621 ANK 270.
FT REPEAT 12622 12662 ANK 271.
FT REPEAT 12663 12703 ANK 272.
FT REPEAT 12704 12744 ANK 273.
FT REPEAT 12745 12785 ANK 274.
FT REPEAT 12786 12825 ANK 275.
FT REPEAT 12826 12866 ANK 276.
FT REPEAT 12867 12906 ANK 277.
FT REPEAT 12907 12947 ANK 278.
FT REPEAT 12948 12988 ANK 279.
FT REPEAT 12989 13029 ANK 280.
FT REPEAT 13030 13070 ANK 281.
FT REPEAT 13071 13111 ANK 282.
FT REPEAT 13112 13152 ANK 283.
FT REPEAT 13153 13193 ANK 284.
FT REPEAT 13194 13234 ANK 285.
FT REPEAT 13235 13275 ANK 286.
FT REPEAT 13276 13316 ANK 287.
FT REPEAT 13317 13357 ANK 288.
FT REPEAT 13358 13398 ANK 289.
FT REPEAT 13399 13439 ANK 290.
FT REPEAT 13440 13480 ANK 291.
FT REPEAT 13481 13521 ANK 292.
FT REPEAT 13522 13562 ANK 293.
FT REPEAT 13563 13603 ANK 294.
FT REPEAT 13604 13644 ANK 295.
FT REPEAT 13645 13685 ANK 296.
FT REPEAT 13686 13725 ANK 297.
FT REPEAT 13726 13766 ANK 298.
FT REPEAT 13767 13806 ANK 299.
FT REPEAT 13807 13847 ANK 300.
FT REPEAT 13848 13888 ANK 301.
FT REPEAT 13889 13928 ANK 302.
FT REPEAT 13929 13969 ANK 303.
FT REPEAT 13970 14010 ANK 304.
FT REPEAT 14011 14051 ANK 305.
FT REPEAT 14052 14092 ANK 306.
FT REPEAT 14093 14133 ANK 307.
FT REPEAT 14134 14174 ANK 308.
FT REPEAT 14175 14215 ANK 309.
FT REPEAT 14216 14256 ANK 310.
FT REPEAT 14257 14296 ANK 311.
FT REPEAT 14297 14337 ANK 312.
FT REPEAT 14338 14378 ANK 313.
FT REPEAT 14379 14418 ANK 314.
FT REPEAT 14419 14459 ANK 315.
FT REPEAT 14460 14500 ANK 316.
FT REPEAT 14501 14541 ANK 317.
FT REPEAT 14542 14582 ANK 318.
FT REPEAT 14583 14623 ANK 319.
FT REPEAT 14624 14664 ANK 320.
FT REPEAT 14665 14704 ANK 321.
FT REPEAT 14705 14745 ANK 322.
FT REPEAT 14746 14786 ANK 323.
FT REPEAT 14787 14826 ANK 324.
FT REPEAT 14827 14867 ANK 325.
FT REPEAT 14868 14907 ANK 326.
FT REPEAT 14908 14948 ANK 327.
FT REPEAT 14949 14989 ANK 328.
FT REPEAT 14990 15029 ANK 329.
FT REPEAT 15030 15070 ANK 330.
FT REPEAT 15071 15111 ANK 331.
FT REPEAT 15112 15152 ANK 332.
FT REPEAT 15153 15193 ANK 333.
FT REPEAT 15194 15234 ANK 334.
FT REPEAT 15235 15275 ANK 335.
FT REPEAT 15276 15316 ANK 336.
FT REPEAT 15317 15357 ANK 337.
FT REPEAT 15358 15398 ANK 338.
FT REPEAT 15399 15439 ANK 339.
FT REPEAT 15440 15480 ANK 340.
FT REPEAT 15481 15521 ANK 341.
FT REPEAT 15522 15562 ANK 342.
FT REPEAT 15563 15603 ANK 343.
FT REPEAT 15604 15644 ANK 344.
FT REPEAT 15645 15685 ANK 345.
FT REPEAT 15686 15725 ANK 346.
FT REPEAT 15726 15766 ANK 347.
FT REPEAT 15767 15806 ANK 348.
FT REPEAT 15807 15847 ANK 349.
FT REPEAT 15848 15888 ANK 350.
FT REPEAT 15889 15928 ANK 351.
FT REPEAT 15929 15969 ANK 352.
FT REPEAT 15970 16010 ANK 353.
FT REPEAT 16011 16051 ANK 354.
FT REPEAT 16052 16092 ANK 355.
FT REPEAT 16093 16133 ANK 356.
FT REPEAT 16134 16174 ANK 357.
FT REPEAT 16175 16215 ANK 358.
FT REPEAT 16216 16256 ANK 359.
FT REPEAT 16257 16296 ANK 360.
FT REPEAT 16297 16337 ANK 361.
FT REPEAT 16338 16378 ANK 362.
FT REPEAT 16379 16418 ANK 363.
FT REPEAT 16419 16459 ANK 364.
FT REPEAT 16460 16500 ANK 365.
FT REPEAT 16501 16541 ANK 366.
FT REPEAT 16542 16582 ANK 367.
FT REPEAT 16583 16623 ANK 368.
FT REPEAT 16624 16664 ANK 369.
FT REPEAT 16665 16704 ANK 370.
FT REPEAT 16705 16745 ANK 371.
FT REPEAT 16746 16785 ANK 372.
FT REPEAT 16786 16826 ANK 373.
FT REPEAT 16827 16866 ANK 374.
FT REPEAT 16867 16907 ANK 375.
FT REPEAT 16908 16948 ANK 376.
FT REPEAT 16949 16989 ANK 377.
FT REPEAT 16990 17029 ANK 378.
FT REPEAT 17030 17070 ANK 379.
FT REPEAT 17071 17111 ANK 380.
FT REPEAT 17112 17152 ANK 381.
FT REPEAT 17153 17193 ANK 382.
FT REPEAT 17194 17234 ANK 383.
FT REPEAT 17235 17275 ANK 384.
FT REPEAT 17276 17316 ANK 385.
FT REPEAT 17317 17357 ANK 386.
FT REPEAT 17358 17398 ANK 387.
FT REPEAT 17399 17439 ANK 388.
FT REPEAT 17440 17480 ANK 389.
FT REPEAT 17481 17521 ANK 390.
FT REPEAT 17522 17562 ANK 391.
FT REPEAT 17563 17603 ANK 392.
FT REPEAT 17604 17644 ANK 393.
FT REPEAT 17645 17685 ANK 394.
FT REPEAT 17686 17725 ANK 395.
FT REPEAT 17726 17766 ANK 396.
FT REPEAT 17767 17806 ANK 397.
FT REPEAT 17807 17847 ANK 398.
FT REPEAT 17848 17888 ANK 399.
FT REPEAT 17889 17928 ANK 400.
FT REPEAT 17929 17969 ANK 401.
FT REPEAT 17970 18010 ANK 402.
FT REPEAT 18011 18051 ANK 403.
FT REPEAT 18052 18092 ANK 404.
FT REPEAT 18093 18133 ANK 405.
FT REPEAT 18134 18174 ANK 406.
FT REPEAT 18175 18215 ANK 407.
FT REPEAT 18216 18256 ANK 408.
FT REPEAT 18257 18296 ANK 409.
FT REPEAT 18297 18337 ANK 410.
FT REPEAT 18338 18378 ANK 411.
FT REPEAT 18379 18418 ANK 412.
FT REPEAT 18419 18459 ANK 413.
FT REPEAT 18460 18500 ANK 414.
FT REPEAT 18501 18541 ANK 415.
FT REPEAT 18542 18582 ANK 416.
FT REPEAT 18583 18623 ANK 417.
FT REPEAT 18624 18664 ANK 418.
FT REPEAT 18665 18704 ANK 419.
FT REPEAT 18705 18745 ANK 420.
FT REPEAT 18746 18785 ANK 421.
FT REPEAT 18786 18826 ANK 422.
FT REPEAT 18827 18866 ANK 423.
FT REPEAT 18867 18907 ANK 424.
FT REPEAT 18908 18948 ANK 425.
FT REPEAT 18949 18989 ANK 426.
FT REPEAT 18990 19029 ANK 427.
FT REPEAT 19030 19070 ANK 428.
FT REPEAT 19071 19111 ANK 429.
FT REPEAT 19112 19152 ANK 430.
FT REPEAT 19153 19193 ANK 431.
FT REPEAT 19194 19234 ANK 432.
FT REPEAT 19235 19275 ANK 433.
FT REPEAT 19276 19316 ANK 434.
FT REPEAT 19317 19357 ANK 435.
FT REPEAT 19358 19398 ANK 436.
FT REPEAT 19399 19439 ANK 437.
FT REPEAT 19440 19480 ANK 438.
FT REPEAT 19481 19521 ANK 439.
FT REPEAT 19522 19562 ANK 440.
FT REPEAT 19563 19603 ANK 441.
FT REPEAT 19604 19644 ANK 442.
FT REPEAT 19645 19685 ANK 443.
FT REPEAT 19686 19725 ANK 444.
FT REPEAT 19726 19766 ANK 445.
FT REPEAT 19767 19806 ANK 446.
FT REPEAT 19807 19847 ANK 447.
FT REPEAT 19848 19888 ANK 448.
FT REPEAT 19889 19928 ANK 449.
FT REPEAT 19929 19969 ANK 450.
FT REPEAT 19970 20010 ANK 451.
FT REPEAT 20011 20051 ANK 452.
FT REPEAT 20052 20092 ANK 453.
FT REPEAT 20093 20133 ANK 454.
FT REPEAT 20134 20174 ANK 455.
FT REPEAT 20175 20215 ANK 456.
FT REPEAT 20216 20256 ANK 457.
FT REPEAT 20257 20296 ANK 458.
FT REPEAT 20297 20337 ANK 459.
FT REPEAT 20338 20378 ANK 460.
FT REPEAT 20379 20418 ANK 461.
FT REPEAT 20419 20459 ANK 462.
FT REPEAT 20460 20500 ANK 463.
FT REPEAT 20501 20541 ANK 464.
FT REPEAT 20542 20582 ANK 465.
FT REPEAT 20583 20623 ANK 466.
FT REPEAT 20624 20664 ANK 467.
FT REPEAT 20665 20704 ANK 468.
FT REPEAT 20705 20745 ANK 469.
FT REPEAT 20746 20785 ANK 470.
FT REPEAT 20786 20826 ANK 471.
FT REPEAT 20827 20866 ANK 472.
FT REPEAT 20867 20907 ANK 473.
FT REPEAT 20908 20948 ANK 474.
FT REPEAT 20949 20989 ANK 475.
FT REPEAT 20990 21029 ANK 476.
FT REPEAT 21030 21070 ANK 477.
FT REPEAT 21071 21111 ANK 478.
FT REPEAT 21112 21152 ANK 479.
FT REPEAT 21153 21193 ANK 480.
FT REPEAT 21194 21234 ANK 481.
FT REPEAT 21235 21275 ANK 482.
FT REPEAT 21276 21316 ANK 483.
FT REPEAT 21317 21357 ANK 484.
FT REPEAT 21358 21398 ANK 485.
FT REPEAT 21399 21439 ANK 486.
FT REPEAT 21440 21480 ANK 487.
FT REPEAT 21481 21521 ANK 488.
FT REPEAT 21522 21562 ANK 489.
FT REPEAT 21563 21603 ANK 490.
FT REPEAT 21604 21644 ANK 491.
FT REPEAT 21645 21685 ANK 492.
FT REPEAT 21686 21725 ANK 493.
FT REPEAT 21726 21766 ANK 494.
FT REPEAT 21767 21806 ANK 495.
FT REPEAT 21807 21847 ANK 496.
FT REPEAT 21848 21888 ANK 497.
FT REPEAT 21889 21928 ANK 498.
FT REPEAT 21929 21969 ANK 499.
FT REPEAT 21970 22010 ANK 500.
FT REPEAT 22011 22051 ANK 501.
FT REPEAT 22052 22092 ANK 502.
FT REPEAT 22093 22133 ANK 503.
FT REPEAT 22134 22174 ANK 504.
FT REPEAT 22175 22215 ANK 505.
FT REPEAT 22216 22256 ANK 506.
FT REPEAT 22257 22296 ANK 507.
FT REPEAT 22297 22337 ANK 508.
FT REPEAT 22338 22378 ANK 509.
FT REPEAT 22379 22418 ANK 510.
FT REPEAT 22419 22459 ANK 511.
FT REPEAT 22460 22500 ANK 512.
FT REPEAT 22501 22541 ANK 513.
FT REPEAT 22542 22582 ANK 514.
FT REPEAT 22583 22623 ANK 515.
FT REPEAT 22624 22664 ANK 516.
FT REPEAT 22665 22704 ANK 517.
FT REPEAT 22705 22745 ANK 518.
FT REPEAT 22746 22785 ANK 519.
FT REPEAT 22786 22826 ANK 520.
FT REPEAT 22827 22866 ANK 521.
FT REPEAT 22867 22907 ANK 522.
FT REPEAT 22908 22948 ANK 523.
FT REPEAT 22949 22989 ANK 524.
FT REPEAT 22990 23029 ANK 525.
FT REPEAT 23030 23070 ANK 526.
FT REPEAT 23071 23111 ANK 527.
FT REPEAT 23112 23152 ANK 528.
FT REPEAT 23153 23193 ANK 529.
FT REPEAT 23194 23
```

Db 1054 COSLVKMCSSPCKNGSGCMQOGASFTCCQASGNTGICDVPVSCEVAARQGVAVL 1113
 QY 443 -----CPLGTGYGNCSSRCG-----CKNDAYCSPVD--SCTCKAGN 477
 Db 1114 CRHAGCCVDAGNTHLCKCAQGYGYSQEDVDEQPNPCNGARTCTYLGYSCECPGX 1173
 QY 478 HGVDGS-----INCPSTGWFGCNL--TC-----499
 Db 1174 HGMKCSKEINECLSQPCQNGCTCIDLVNTYKSCSPRGTQVHCIEDIDCSFVPLTGE 1233
 QY 500 -OCLNGACNTLDG--TCTCAPGMRGEKE-----LPCQ-DGTGLMCAE-----RC 542
 Db 1234 PCPENGRCVDRVGIGCVCPAGVGRGCEGDVNECLSDPCDPSG--NCVQLINDRC 1291
 QY 543 DCSHA-----DGHPT-----TGH--CRLPGSGVHCD-----569
 Db 1292 ECRTGYGRKCEYVENCCKPTCKNGCTCAVASTKRGYICKQPRGSSSCETDSQSG 1351
 QY 570 -----SVCAGRMGPN 581
 Db 1352 SLRCRNGATCVSGHLSPRC 1370

RESULT 4
 ID NTG3_HUMAN STANDARD; PRT; 2321 AA.
 AC Q9UM47; Q9Y6L8; Q9UPL3; Q9UEB3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE-97032728; PubMed-8878478;
 RA Joutel A., Corpechot C., Ducros A., Vahedi K., Chabrier H., Mouton P.,
 RA Alamowitch S., Domenga V., Cecillion M., Marchal E., MacLazek J.,
 RA Vaysiere C., Cruaud C., Cabanis E.-A., Ruchoux M.M., Weissbach J.,
 RA Bach J.-F., Bousset M.-G., Tournier-Lasserre E.,
 RT "Notch3 mutations in CADASIL, a hereditary adult-onset condition
 causing stroke and dementia.";
 RL Nature 383:707-710(1996).
 RN [2]
 RX SEQUENCE FROM N.A.
 RA Gunel M., Artavanis-Tsakonas S.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RX SEQUENCE FROM N.A.
 RA Lemerlin J.E., McCreedy P.M., Skowronski E., Adams A.M.,
 RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilaugen S.,
 RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
 RA Christensen M., Georgescu A., Avila J., Liu S., Attkin A., Andreise T.,
 RA Traubheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
 RA Bruce R., Thomas P., Quan G., Krommiller B., Ariellano A.,
 RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
 RA Carraro A.V.;
 RT "Sequence analysis of an 1.5 Mb olfactory receptor (OLFR) cluster in
 19p13.1";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RX VARIANTS CADASIL.Y-49; C-71; C-90; C-110; C-133; C-141; R-146; C-153;
 RP C-169; C-171; C-182; R-185; S-212; G-222; Y-224; C-258; Y-542; C-558;
 RP C-578; C-728; C-985; C-1006; C-1031; C-1231 AND R-1261, AND VARIANTS
 RP R-170; L-496; Q-1133; M-1183 AND A-2223.
 RX MEDLINE-98049753; PubMed-9388399;
 RA Joutel A., Vahedi K., Corpechot C., Troesch A., Chabrier H.,
 RA Vaysiere C., Cruaud C., MacLazek J., Weissbach J., Bousset M.-G.,
 RA Bach J.-F., Tournier-Lasserre E.;
 RT "Strong clustering and stereotyped nature of Notch3 mutations in

RT CADASIL patients.";
 RL Lancet 350:1511-1515(1997).
 RN [5]
 RP VARIANT CADASIL.114-GLY--PRO-120 DEL.
 RX MEDLINE-20264473; PubMed-10802807;
 RA Joutel A., Chabrier H., Vahedi K., Domenga V., Vaysiere C.,
 RA Ruchoux M.M., Lucas C., Leys D., Bousset M.-G., Tournier-Lasserre E.;
 RT "Splice site mutation causing a seven amino acid Notch3 in-frame
 deletion in CADASIL.";
 RL Neurology 54:1874-1875(2000).
 RN [6]
 RP IDENTIFICATION OF LIGANDS.
 RX MEDLINE-99180765; PubMed-10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed in fetal and adult
 CC tissues.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- DISEASE: Defects in NOTCH3 are associated with CADASIL (Cerebral
 CC autosomal dominant arteriopathy with subcortical infarcts and
 CC leukoencephalopathy) which causes a type of stroke and dementia of
 CC which key features include recurrent subcortical ischemic events
 CC and vascular dementia.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; U97669; AAB91371.1; -
 DR EMBL; AF058900; AAC14346.1; -
 DR EMBL; AF058881; AAC14346.1; JOINED.
 DR EMBL; AF058882; AAC14346.1; JOINED.
 DR EMBL; AF058883; AAC14346.1; JOINED.
 DR EMBL; AF058884; AAC14346.1; JOINED.
 DR EMBL; AF058885; AAC14346.1; JOINED.
 DR EMBL; AF058886; AAC14346.1; JOINED.
 DR EMBL; AF058887; AAC14346.1; JOINED.
 DR EMBL; AF058888; AAC14346.1; JOINED.
 DR EMBL; AF058889; AAC14346.1; JOINED.
 DR EMBL; AF058890; AAC14346.1; JOINED.

DR EMBL; AF058891; AAC14346.1; JOINED.
 DR EMBL; AF058892; AAC14346.1; JOINED.
 DR EMBL; AF058893; AAC14346.1; JOINED.
 DR EMBL; AF058894; AAC14346.1; JOINED.
 DR EMBL; AF058895; AAC14346.1; JOINED.
 DR EMBL; AF058896; AAC14346.1; JOINED.
 DR EMBL; AF058897; AAC14346.1; JOINED.
 DR EMBL; AF058898; AAC14346.1; JOINED.
 DR EMBL; AF058899; AAC14346.1; JOINED.
 DR EMBL; AC004257; AAC04897.1; JOINED.
 DR EMBL; AC004663; AAC15789.1; ALT_INIT.
 DR HSSP; P00740; 1EDM.
 DR Genew; HGNC:7883; NOTCH3.
 DR MIM; 600276; -.
 DR MIM; 125310; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; ASX_hydroxyl.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR000742; EGF-2.
 DR InterPro; IPR001881; EGF_Ca.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR002049; Laminin_EGF.
 DR InterPro; IPR000800; Notch.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00066; notch; 3.
 DR PRINTS; PR00010; EGFBL00D.
 DR PRINTS; PR00011; EGFAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 5.
 DR SMART; SM00179; EGF_CA; 19.
 DR SMART; SM00001; EGF_Like; 14.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 18.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 25.
 DR PROSITE; PS01187; EGF_CA; 18.
 KW Receptor: Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Disease mutation.
 FT SIGNAL 1 39 POTENTIAL.
 FT CHAIN 40 2321 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 3.
 FT CHAIN 1629 2321 NOTCH EXTRACELLULAR TRUNCATION (BY
 FT CHAIN 1662 2321 SIMILARITY).
 FT CHAIN 1644 1643 NOTCH INTRACELLULAR DOMAIN (BY
 FT TRANSMEM 1644 1664 EXTRACELLULAR (POTENTIAL).
 FT 1665 2321 POTENTIAL.
 FT 40 77 CYTOPLASMIC (POTENTIAL).
 FT 78 118 EGF-Like 1.
 FT 119 156 EGF-Like 2.
 FT 158 195 EGF-Like 3.
 FT 197 234 EGF-Like 4.
 FT 236 272 EGF-Like 5.
 FT 274 312 EGF-Like 6.
 FT 312 350 EGF-Like 7.
 FT 350 389 EGF-Like 8.
 FT 389 429 EGF-Like 9.
 FT 429 467 EGF-Like 10.
 FT 467 505 EGF-Like 11.
 FT 505 543 EGF-Like 12.
 FT 543 580 EGF-Like 13.
 FT 580 618 EGF-Like 14.
 FT 618 655 EGF-Like 15.
 FT 655 693 EGF-Like 16.
 FT 693 730 EGF-Like 17.
 FT 730 770 EGF-Like 18.
 FT 770 808 EGF-Like 19.
 FT 808 847 EGF-Like 20.
 FT 847 887 EGF-Like 21.
 FT 887 922 EGF-Like 22.
 FT 922 960 EGF-Like 23.
 FT 960 998 EGF-Like 24.
 FT 998 1034 EGF-Like 25.
 FT 1034 1082 EGF-Like 26.
 FT 1082 1118 EGF-Like 27.
 FT 1118 1154 EGF-Like 28.
 FT 1154 1191 EGF-Like 29.
 FT 1191 1228 EGF-Like 30.
 FT 1228 1265 EGF-Like 31.
 FT 1265 1302 EGF-Like 32.
 FT 1302 1339 EGF-Like 33.
 FT 1339 1376 EGF-Like 34.
 FT 1376 1413 EGF-Like 35.
 FT 1413 1450 EGF-Like 36.
 FT 1450 1487 EGF-Like 37.
 FT 1487 1524 EGF-Like 38.
 FT 1524 1561 EGF-Like 39.
 FT 1561 1598 EGF-Like 40.
 FT 1598 1635 EGF-Like 41.
 FT 1635 1672 EGF-Like 42.
 FT 1672 1709 EGF-Like 43.
 FT 1709 1746 EGF-Like 44.
 FT 1746 1783 EGF-Like 45.
 FT 1783 1820 EGF-Like 46.
 FT 1820 1857 EGF-Like 47.
 FT 1857 1894 EGF-Like 48.
 FT 1894 1931 EGF-Like 49.
 FT 1931 1968 EGF-Like 50.
 FT 1968 2005 EGF-Like 51.
 FT 2005 2042 EGF-Like 52.
 FT 2042 2079 EGF-Like 53.
 FT 2079 2116 EGF-Like 54.
 FT 2116 2153 EGF-Like 55.
 FT 2153 2190 EGF-Like 56.
 FT 2190 2227 EGF-Like 57.
 FT 2227 2264 EGF-Like 58.
 FT 2264 2301 EGF-Like 59.
 FT 2301 2338 EGF-Like 60.
 FT 2338 2375 EGF-Like 61.
 FT 2375 2412 EGF-Like 62.
 FT 2412 2449 EGF-Like 63.
 FT 2449 2486 EGF-Like 64.
 FT 2486 2523 EGF-Like 65.
 FT 2523 2560 EGF-Like 66.
 FT 2560 2597 EGF-Like 67.
 FT 2597 2634 EGF-Like 68.
 FT 2634 2671 EGF-Like 69.
 FT 2671 2708 EGF-Like 70.
 FT 2708 2745 EGF-Like 71.
 FT 2745 2782 EGF-Like 72.
 FT 2782 2819 EGF-Like 73.
 FT 2819 2856 EGF-Like 74.
 FT 2856 2893 EGF-Like 75.
 FT 2893 2930 EGF-Like 76.
 FT 2930 2967 EGF-Like 77.
 FT 2967 3004 EGF-Like 78.
 FT 3004 3041 EGF-Like 79.
 FT 3041 3078 EGF-Like 80.
 FT 3078 3115 EGF-Like 81.
 FT 3115 3152 EGF-Like 82.
 FT 3152 3189 EGF-Like 83.
 FT 3189 3226 EGF-Like 84.
 FT 3226 3263 EGF-Like 85.
 FT 3263 3300 EGF-Like 86.
 FT 3300 3337 EGF-Like 87.
 FT 3337 3374 EGF-Like 88.
 FT 3374 3411 EGF-Like 89.
 FT 3411 3448 EGF-Like 90.
 FT 3448 3485 EGF-Like 91.
 FT 3485 3522 EGF-Like 92.
 FT 3522 3559 EGF-Like 93.
 FT 3559 3596 EGF-Like 94.
 FT 3596 3633 EGF-Like 95.
 FT 3633 3670 EGF-Like 96.
 FT 3670 3707 EGF-Like 97.
 FT 3707 3744 EGF-Like 98.
 FT 3744 3781 EGF-Like 99.
 FT 3781 3818 EGF-Like 100.
 FT 3818 3855 EGF-Like 101.
 FT 3855 3892 EGF-Like 102.
 FT 3892 3929 EGF-Like 103.
 FT 3929 3966 EGF-Like 104.
 FT 3966 4003 EGF-Like 105.
 FT 4003 4040 EGF-Like 106.
 FT 4040 4077 EGF-Like 107.
 FT 4077 4114 EGF-Like 108.
 FT 4114 4151 EGF-Like 109.
 FT 4151 4188 EGF-Like 110.
 FT 4188 4225 EGF-Like 111.
 FT 4225 4262 EGF-Like 112.
 FT 4262 4299 EGF-Like 113.
 FT 4299 4336 EGF-Like 114.
 FT 4336 4373 EGF-Like 115.
 FT 4373 4410 EGF-Like 116.
 FT 4410 4447 EGF-Like 117.
 FT 4447 4484 EGF-Like 118.
 FT 4484 4521 EGF-Like 119.
 FT 4521 4558 EGF-Like 120.
 FT 4558 4595 EGF-Like 121.
 FT 4595 4632 EGF-Like 122.
 FT 4632 4669 EGF-Like 123.
 FT 4669 4706 EGF-Like 124.
 FT 4706 4743 EGF-Like 125.
 FT 4743 4780 EGF-Like 126.
 FT 4780 4817 EGF-Like 127.
 FT 4817 4854 EGF-Like 128.
 FT 4854 4891 EGF-Like 129.
 FT 4891 4928 EGF-Like 130.
 FT 4928 4965 EGF-Like 131.
 FT 4965 5002 EGF-Like 132.
 FT 5002 5039 EGF-Like 133.
 FT 5039 5076 EGF-Like 134.
 FT 5076 5113 EGF-Like 135.
 FT 5113 5150 EGF-Like 136.
 FT 5150 5187 EGF-Like 137.
 FT 5187 5224 EGF-Like 138.
 FT 5224 5261 EGF-Like 139.
 FT 5261 5298 EGF-Like 140.
 FT 5298 5335 EGF-Like 141.
 FT 5335 5372 EGF-Like 142.
 FT 5372 5409 EGF-Like 143.
 FT 5409 5446 EGF-Like 144.
 FT 5446 5483 EGF-Like 145.
 FT 5483 5520 EGF-Like 146.
 FT 5520 5557 EGF-Like 147.
 FT 5557 5594 EGF-Like 148.
 FT 5594 5631 EGF-Like 149.
 FT 5631 5668 EGF-Like 150.
 FT 5668 5705 EGF-Like 151.
 FT 5705 5742 EGF-Like 152.
 FT 5742 5779 EGF-Like 153.
 FT 5779 5816 EGF-Like 154.
 FT 5816 5853 EGF-Like 155.
 FT 5853 5890 EGF-Like 156.
 FT 5890 5927 EGF-Like 157.
 FT 5927 5964 EGF-Like 158.
 FT 5964 6001 EGF-Like 159.
 FT 6001 6038 EGF-Like 160.
 FT 6038 6075 EGF-Like 161.
 FT 6075 6112 EGF-Like 162.
 FT 6112 6149 EGF-Like 163.
 FT 6149 6186 EGF-Like 164.
 FT 6186 6223 EGF-Like 165.
 FT 6223 6260 EGF-Like 166.
 FT 6260 6297 EGF-Like 167.
 FT 6297 6334 EGF-Like 168.
 FT 6334 6371 EGF-Like 169.
 FT 6371 6408 EGF-Like 170.
 FT 6408 6445 EGF-Like 171.
 FT 6445 6482 EGF-Like 172.
 FT 6482 6519 EGF-Like 173.
 FT 6519 6556 EGF-Like 174.
 FT 6556 6593 EGF-Like 175.
 FT 6593 6630 EGF-Like 176.
 FT 6630 6667 EGF-Like 177.
 FT 6667 6704 EGF-Like 178.
 FT 6704 6741 EGF-Like 179.
 FT 6741 6778 EGF-Like 180.
 FT 6778 6815 EGF-Like 181.
 FT 6815 6852 EGF-Like 182.
 FT 6852 6889 EGF-Like 183.
 FT 6889 6926 EGF-Like 184.
 FT 6926 6963 EGF-Like 185.
 FT 6963 7000 EGF-Like 186.
 FT 7000 7037 EGF-Like 187.
 FT 7037 7074 EGF-Like 188.
 FT 7074 7111 EGF-Like 189.
 FT 7111 7148 EGF-Like 190.
 FT 7148 7185 EGF-Like 191.
 FT 7185 7222 EGF-Like 192.
 FT 7222 7259 EGF-Like 193.
 FT 7259 7296 EGF-Like 194.
 FT 7296 7333 EGF-Like 195.
 FT 7333 7370 EGF-Like 196.
 FT 7370 7407 EGF-Like 197.
 FT 7407 7444 EGF-Like 198.
 FT 7444 7481 EGF-Like 199.
 FT 7481 7518 EGF-Like 200.
 FT 7518 7555 EGF-Like 201.
 FT 7555 7592 EGF-Like 202.
 FT 7592 7629 EGF-Like 203.
 FT 7629 7666 EGF-Like 204.
 FT 7666 7703 EGF-Like 205.
 FT 7703 7740 EGF-Like 206.
 FT 7740 7777 EGF-Like 207.
 FT 7777 7814 EGF-Like 208.
 FT 7814 7851 EGF-Like 209.
 FT 7851 7888 EGF-Like 210.
 FT 7888 7925 EGF-Like 211.
 FT 7925 7962 EGF-Like 212.
 FT 7962 8000 EGF-Like 213.
 FT 8000 8037 EGF-Like 214.
 FT 8037 8074 EGF-Like 215.
 FT 8074 8111 EGF-Like 216.
 FT 8111 8148 EGF-Like 217.
 FT 8148 8185 EGF-Like 218.
 FT 8185 8222 EGF-Like 219.
 FT 8222 8259 EGF-Like 220.
 FT 8259 8296 EGF-Like 221.
 FT 8296 8333 EGF-Like 222.
 FT 8333 8370 EGF-Like 223.
 FT 8370 8407 EGF-Like 224.
 FT 8407 8444 EGF-Like 225.
 FT 8444 8481 EGF-Like 226.
 FT 8481 8518 EGF-Like 227.
 FT 8518 8555 EGF-Like 228.
 FT 8555 8592 EGF-Like 229.
 FT 8592 8629 EGF-Like 230.
 FT 8629 8666 EGF-Like 231.
 FT 8666 8703 EGF-Like 232.
 FT 8703 8740 EGF-Like 233.
 FT 8740 8777 EGF-Like 234.
 FT 8777 8814 EGF-Like 235.
 FT 8814 8851 EGF-Like 236.
 FT 8851 8888 EGF-Like 237.
 FT 8888 8925 EGF-Like 238.
 FT 8925 8962 EGF-Like 239.
 FT 8962 9000 EGF-Like 240.
 FT 9000 9037 EGF-Like 241.
 FT 9037 9074 EGF-Like 242.
 FT 9074 9111 EGF-Like 243.
 FT 9111 9148 EGF-Like 244.
 FT 9148 9185 EGF-Like 245.
 FT 9185 9222 EGF-Like 246.
 FT 9222 9259 EGF-Like 247.
 FT 9259 9296 EGF-Like 248.
 FT 9296 9333 EGF-Like 249.
 FT 9333 9370 EGF-Like 250.
 FT 9370 9407 EGF-Like 251.
 FT 9407 9444 EGF-Like 252.
 FT 9444 9481 EGF-Like 253.
 FT 9481 9518 EGF-Like 254.
 FT 9518 9555 EGF-Like 255.
 FT 9555 9592 EGF-Like 256.
 FT 9592 9629 EGF-Like 257.
 FT 9629 9666 EGF-Like 258.
 FT 9666 9703 EGF-Like 259.
 FT 9703 9740 EGF-Like 260.
 FT 9740 9777 EGF-Like 261.
 FT 9777 9814 EGF-Like 262.
 FT 9814 9851 EGF-Like 263.
 FT 9851 9888 EGF-Like 264.
 FT 9888 9925 EGF-Like 265.
 FT 9925 9962 EGF-Like 266.
 FT 9962 10000 EGF-Like 267.
 FT 10000 10037 EGF-Like 268.
 FT 10037 10074 EGF-Like 269.
 FT 10074 10111 EGF-Like 270.
 FT 10111 10148 EGF-Like 271.
 FT 10148 10185 EGF-Like 272.
 FT 10185 10222 EGF-Like 273.
 FT 10222 10259 EGF-Like 274.
 FT 10259 10296 EGF-Like 275.
 FT 10296 10333 EGF-Like 276.
 FT 10333 10370 EGF-Like 277.
 FT 10370 10407 EGF-Like 278.
 FT 10407 10444 EGF-Like 279.
 FT 10444 10481 EGF-Like 280.
 FT 10481 10518 EGF-Like 281.
 FT 10518 10555 EGF-Like 282.
 FT 10555 10592 EGF-Like 283.
 FT 10592 10629 EGF-Like 284.
 FT 10629 10666 EGF-Like 285.
 FT 10666 10703 EGF-Like 286.
 FT 10703 10740 EGF-Like 287.
 FT 10740 10777 EGF-Like 288.
 FT 10777 10814 EGF-Like 289.
 FT 10814 10851 EGF-Like 290.
 FT 10851 10888 EGF-Like 291.
 FT 10888 10925 EGF-Like 292.
 FT 10925 10962 EGF-Like 293.
 FT 10962 11000 EGF-Like 294.
 FT 11000 11037 EGF-Like 295.
 FT 11037 11074 EGF-Like 296.
 FT 11074 11111 EGF-Like 297.
 FT 11111 11148 EGF-Like 298.
 FT 11148 11185 EGF-Like 299.
 FT 11185 11222 EGF-Like 300.
 FT 11222 11259 EGF-Like 301.
 FT 11259 11296 EGF-Like 302.
 FT 11296 11333 EGF-Like 303.
 FT 11333 11370 EGF-Like 304.
 FT 11370 11407 EGF-Like 305.
 FT 11407 11444 EGF-Like 306.
 FT 11444 11481 EGF-Like 307.
 FT 11481 11518 EGF-Like 308.
 FT 11518 11555 EGF-Like 309.
 FT 11555 11592 EGF-Like 310.
 FT 11592 11629 EGF-Like 311.
 FT 11629 11666 EGF-Like 312.
 FT 11666 11703 EGF-Like 313.
 FT 11703 11740 EGF-Like 314.
 FT 11740 11777 EGF-Like 315.
 FT 11777 11814 EGF-Like 316.
 FT 11814 11851 EGF-Like 317.
 FT 11851 11888 EGF-Like 318.
 FT 11888 11925 EGF-Like 319.
 FT 11925 11962 EGF-Like 320.
 FT 11962 12000 EGF-Like 321.
 FT 12000 12037 EGF-Like 322.
 FT 12037 12074 EGF-Like 323.
 FT 12074 12111 EGF-Like 324.
 FT 12111 12148 EGF-Like 325.
 FT 12148 12185 EGF-Like 326.
 FT 12185 12222 EGF-Like 327.
 FT 12222 12259 EGF-Like 328.
 FT 12259 12296 EGF-Like 329.
 FT 12296 12333 EGF-Like 330.
 FT 12333 12370 EGF-Like 331.
 FT 12370 12407 EGF-Like 332.
 FT 12407 12444 EGF-Like 333.
 FT 12444 12481 EGF-Like 334.
 FT 12481 12518 EGF-Like 335.
 FT 12518 12555 EGF-Like 336.
 FT 12555 12592 EGF-Like 337.
 FT 12592 12629 EGF-Like 338.
 FT 12629 12666 EGF-Like 339.
 FT 12666 12703 EGF-Like 340.
 FT 12703 12740 EGF-Like 341.
 FT 12740 12777 EGF-Like 342.
 FT 12777 12814 EGF-Like 343.
 FT 12814 12851 EGF-Like 344.
 FT 12851 12888 EGF-Like 345.
 FT 12888 12925 EGF-Like 346.
 FT 12925 12962 EGF-Like 347.
 FT 12962 13000 EGF-Like 348.
 FT 13000 13037 EGF-Like 349.
 FT 13037 13074 EGF-Like 350.
 FT 13074 13111 EGF-Like 351.
 FT 13111 13148 EGF-Like 352.
 FT 13148 13185 EGF-Like 353.
 FT 13185 13222 EGF-Like 354.
 FT 13222 13259 EGF-Like 355.
 FT 13259 13296 EGF-Like 356.
 FT 13296 13333 EGF-Like 357.
 FT 13333 13370 EGF-Like 358.
 FT 13370 13407 EGF-Like 359.
 FT 13407 13444 EGF-Like 360.
 FT 13444 13481 EGF-Like 361.
 FT 13481 13518 EGF-Like 362.
 FT 13518 13555 EGF-Like 363.
 FT 13555 13592 EGF-Like 364.
 FT 13592 13629 EGF-Like 365.
 FT 13629 13666 EGF-Like 366.
 FT 13666 13703 EGF-Like 367.
 FT 13703 13740 EGF-Like 368.
 FT 13740 13777 EGF-Like 369.
 FT 13777 13814 EGF-Like 370.
 FT 13814 13851 EGF-Like 371.
 FT 13851 13888 EGF-Like 372.
 FT 13888 13925 EGF-Like 373.
 FT 13925 13962 EGF-Like 374.
 FT 13962 14000 EGF-Like 375.
 FT 14000 14037 EGF-Like 376.
 FT 14037 14074 EGF-Like 377.
 FT 14074 14111 EGF-Like 378.
 FT 14111 14148 EGF-Like 379.
 FT 14148 14185 EGF-Like 380.
 FT 14185 14222 EGF-Like 381.
 FT 14222 14259 EGF-Like 382.
 FT 14259 14296 EGF-Like 383.
 FT 14296 14333 EGF-Like 384.
 FT 14333 14370 EGF-Like 385.
 FT 14370 14407 EGF-Like 386.
 FT 14407 14444 EGF-Like 387.
 FT 14444 14481 EGF-Like 388.
 FT 14481 14518 EGF-Like 389.
 FT 14518 14555 EGF-Like 390.
 FT 14555 14592 EGF-Like 391.
 FT 14592 14629 EGF-Like 392.
 FT 14629 14666 EGF-Like 393.
 FT 14666 14703 EGF-Like 394.
 FT 14703 14740 EGF-Like 395.
 FT 14740 14777 EGF-Like 396.
 FT 14777 14814 EGF-Like 397.
 FT 14814 14851 EGF-Like 398.
 FT 14851 14888 EGF-Like 399.
 FT 14888 14925 EGF-Like 400.
 FT 14925 14962 EGF-Like 401.
 FT 14962 15000 EGF-Like 402.
 FT 15000 15037 EGF-Like 403.
 FT 15037 15074 EGF-Like 404.
 FT 15074 15111 EGF-Like 405.
 FT 15111 15148 EGF-Like 406.
 FT 15148 15185 EGF-Like 407.
 FT 15185 15222 EGF-Like 408.
 FT 15222 15259 EGF-Like 409.
 FT 15259 15296 EGF-Like 410.
 FT 15296 15333 EGF-Like 411.
 FT 15333 15370 EGF-Like 412.
 FT 15370 15407 EGF-Like 413.
 FT 15407 15444 EGF-Like 414.
 FT 15444 15481 EGF-Like 415.
 FT 15481 15518 EGF-Like 416.
 FT 15518 15555 EGF-Like 417.
 FT 15555 15592 EGF-Like 418.
 FT 15592 15629 EGF-Like 419.
 FT 15629 15666 EGF-Like 420.
 FT 15666 15703 EGF-Like 421.
 FT 15703 15740 EGF-Like 422.
 FT 15740 15777 EGF-Like 423.
 FT 15777 15814 EGF-Like 424.
 FT 15814 15851 EGF-Like 425.
 FT 15851 15888 EGF-Like 426.
 FT 15888 15925 EGF-Like 427.
 FT 15925 15962 EGF-Like 428.
 FT 15962 16000 EGF-Like 429.
 FT 16000 16037 EGF-Like 430.
 FT 16037 16074 EGF-Like 431.
 FT 16074 16111 EGF-Like 432.
 FT 16111 16148 EGF-Like 433.
 FT 16148 16185 EGF-Like 434.
 FT 16185 16222 EGF-Like 435.
 FT 16222 16259 EGF-Like 436.
 FT 16259 16296 EGF-Like 437.
 FT 16296 16333 EGF-Like 438.
 FT 16333 16370 EGF-Like 439.
 FT 16370 16407 EGF-Like 440.
 FT 16407 16444 EGF-Like 441.
 FT 16444 16481 EGF-Like 442.
 FT 16481 16518 EGF-Like 443.
 FT 16518 16555 EGF-Like 444.
 FT 16555 16592 EGF-Like 445.
 FT 16592 16629 EGF-Like 446.
 FT 16629 16666 EGF-Like 447.
 FT 16666 16703 EGF-Like 448.
 FT 16703 16740 EGF-Like 449.
 FT 16740 16777 EGF-Like 450.
 FT 16777 16814 EGF-Like 451.
 FT 16814 1

NC1_MOUSE STANDARD; PRT: 2531 AA.
 ID NC1_MOUSE 09QW58; 09QJ2; 006007; 061905; 09R0X7;
 AC 001705; 09QW58; 09QJ2; 006007; 061905; 09R0X7;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neuregulin locus notch homolog protein 1 precursor (Notch 1) (Moch A)
 DE (m14) (p300).
 OS NOTCH1 OR NOTCH.
 OS Mus musculus (Mouse).
 CC Eukaryota; Euteleostomi; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE-Embryo;
 RX MEDLINE-93194170; PubMed-8449489;
 RA Franco del Amo F., Gendron-Maguire M., Swiatek P.J., Jenkins N.A.,
 RA Copeland N.G., Gridley T.;
 RT "Cloning, analysis, and chromosomal localization of Notch-1, a mouse
 RT homolog of Drosophila Notch.";
 RL Genomics 15:259-264(1993).
 [2]
 RP SEQUENCE OF 731-1899 FROM N.A. (ISOFORM 2), AND DEVELOPMENTAL STAGE.
 RC STRAIN-CD-1; TISSUE-Embryo;
 RX MEDLINE-93050801; PubMed-1426644;
 RA Reaume A.G., Conlon R.A., Ziringibl R., Yamaguchi T.P., Rossant J.;
 RT "Expression analysis of a Notch homologue in the mouse embryo.";
 RL Dev. Biol. 154:377-387(1992).
 [3]
 RP SEQUENCE OF 1551-1647 FROM N.A. (ISOFORM 1), AND DEVELOPMENTAL STAGE.
 RC TISSUE-Embryo;
 RX MEDLINE-93048835; PubMed-1425352;
 RA Franco del Amo F., Smith D.E., Swiatek P.J., Gendron-Maguire M.,
 RA Greenspan R.J., McMahon A.P., Gridley T.;
 RT "Expression pattern of Notch, a mouse homologue of Drosophila Notch,
 RT suggests an important role in early postimplantation mouse
 RT development.";
 RL Development 115:737-744(1992).
 [4]
 RP SEQUENCE OF 1161-1547 FROM N.A.
 RC STRAIN-C57BL/6 X CBA; TISSUE-Embryo;
 RX MEDLINE-93178563; PubMed-8440332;
 RA Larelli M., Lendahl U.;
 RT "Moch A and Moch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues.";
 RL Exp. Cell Res. 204:364-372(1993).
 [5]
 RP SEQUENCE OF 1659-1673 FROM N.A.
 RX MEDLINE-99364499; PubMed-10437788;
 RA Lee J.S., Ishimoto A., Yanagawa S.T.;
 RT "Murine leukemia provirus-mediated activation of the Notch1 gene leads
 RT to induction of HES-1 in a mouse T lymphoma cell line, DL-3.";
 RL FEBS Lett. 455:276-280(1999).
 [6]
 RP SEQUENCE OF 1950-2201 FROM N.A.
 RX MEDLINE-98029496; PubMed-9384671;
 RA Messierie M., Folio M., Nehls M., Eggert H., Boehm T.;
 RT "Dynamic changes in gene expression during in vitro differentiation of
 RT mouse embryonic stem cells.";
 RL Cytokines Cell. Mol. Ther. 1:139-143(1995).
 [7]
 RP SEQUENCE OF 1655-1659, CLEAVAGE BY FURIN-LIKE CONVERTASE, AND
 RP MUTAGENESIS OF 1651-ARG-GLN-ARG-1654.
 RX MEDLINE-98318619; PubMed-9653148;
 RA Logeat F., Bessia C., Brou C., Leball O., Jarriault S., Seidah N.G.,
 RA Israel A.;
 RT "The Notch1 receptor is cleaved constitutively by a furin-like
 RT convertase.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8108-8112(1998).
 [8]
 RP PARTIAL SEQUENCE, AND POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE-21523956; PubMed-11518718;

RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (NL-4) undergo presenilin-dependent
 RT proteolysis.";
 RL J. Biol. Chem. 276:40268-40273(2001).
 [9]
 RP POST-TRANSLATIONAL PROCESSING.
 RX MEDLINE-21374376; PubMed-11459941;
 RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-1 kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation. May be involved in mesoderm
 CC development, somitogenesis and neurogenesis. Involved in the
 CC maturation of both CD4+ and CD8+ cells in the thymus.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(Tw) and a N-
 CC terminal fragment N(Ec) which are probably linked by disulfide
 CC bonds.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the brain, lung and
 CC thymus. Expressed at lower levels in the spleen, bone-marrow,
 CC spinal cord, eyes, mammary gland, liver, intestine, skeletal
 CC muscle, kidney and heart.
 CC -1- DEVELOPMENTAL STAGE: First detected in the mesoderm at 7.5 d.p.c.
 CC By 8.5 d.p.c. highly expressed in presomitic mesoderm, mesenchyme
 CC and endothelial cells, while much lower levels are seen in the
 CC neuroepithelium. Between 9.5-10.5 d.p.c. expressed at high levels
 CC in the neuroepithelium. At 13.5 d.p.c. expressed in the surface
 CC ectoderm, eye and developing whisker follicles.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(Tw) and a N-terminal fragment N(Ec). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC -1- PTM: Phosphorylated.
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC EMBL: 211886; CAA79941.1; -
 CC EMBL: 102613; AAK14898.1; -
 CC EMBL: X68278; CAA48339.1; -
 CC EMBL: AJ238029; CAB40733.1; -
 CC EMBL: X82562; CAA57909.1; -
 CC HSP: P00740; IEDM.
 CC MGD: MGI:97363; Notch1.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.

FT DISULFID 241 252 BY SIMILARITY.
 FT DISULFID 246 261 BY SIMILARITY.
 FT DISULFID 263 272 BY SIMILARITY.
 FT DISULFID 279 292 BY SIMILARITY.
 FT DISULFID 286 301 BY SIMILARITY.
 FT DISULFID 303 312 BY SIMILARITY.
 FT DISULFID 319 330 BY SIMILARITY.
 FT DISULFID 324 339 BY SIMILARITY.
 FT DISULFID 341 350 BY SIMILARITY.
 FT DISULFID 356 367 BY SIMILARITY.
 FT DISULFID 361 378 BY SIMILARITY.
 FT DISULFID 380 389 BY SIMILARITY.
 FT DISULFID 396 409 BY SIMILARITY.
 FT DISULFID 403 418 BY SIMILARITY.
 FT DISULFID 420 429 BY SIMILARITY.
 FT DISULFID 436 447 BY SIMILARITY.
 FT DISULFID 441 456 BY SIMILARITY.
 FT DISULFID 458 467 BY SIMILARITY.
 FT DISULFID 474 485 BY SIMILARITY.
 FT DISULFID 479 494 BY SIMILARITY.
 FT DISULFID 496 505 BY SIMILARITY.
 FT DISULFID 512 523 BY SIMILARITY.
 FT DISULFID 517 532 BY SIMILARITY.
 FT DISULFID 534 543 BY SIMILARITY.
 FT DISULFID 550 560 BY SIMILARITY.
 FT DISULFID 555 569 BY SIMILARITY.
 FT DISULFID 571 580 BY SIMILARITY.
 FT DISULFID 587 598 BY SIMILARITY.
 FT DISULFID 592 607 BY SIMILARITY.
 FT DISULFID 609 618 BY SIMILARITY.
 FT DISULFID 625 635 BY SIMILARITY.

Query Match 19.08; Score 685.5; DB 1; Length 2318;

Best Local Similarity 24.48; Pred. No. 1.6e-36; Matches 216; Conservative 59; Mismatches 195; Indels 417; Gaps 48;

QY 94 CCGFYEGEMCVPHCADKCVHGRICIAVNT-----CCCEPQMGSTNCSSACDGDHW 144
 DB 225 CLPGE--EGNCEVNA-VDDCGHRLNGTCVDGVTNVCOCPPRMTQFCTEDVD----- 277
 QY 145 GPHCTSRQ-----CKNALCNPTTG--ACHCAAG-----FRGRNCE 179
 DB 278 -----ECQLPQNMCHNGTCFNLGHSVCVNGMTGESCSQNTIDCATAVCFHGATCH 331
 QY 180 DR-----CEGTYGNDCH--QRC--QCONGATCD--HYTGE--GCRPGYTGAPCED 223
 DB 332 DRVASFYCACPMGKTGLLCHLDACVSNPCHEDAICDINPVSGRAICTCPGFTGGACDQ 391
 QY 224 -----LCPGKHGPOCE----- 235
 DB 392 DVDECSIGANPCENHGRVNTGSRFLCCGGRGTGPRRETVDVNECLSGPCRNQATCLDRI 451
 QY 236 -----QRCPCQNGVC--HHVTG--ECSCPSGMGTVC----- 264
 DB 452 GQFTICMAGFTGYCEVDIDECOSSPCVNGVCKDRVNGFSCITCPSGSGSMQLDVDE 511
 QY 265 -----GPR-----CPEGRGKNCQ--ECQ--CHNGTCTCA--ATGQCHC 298
 DB 512 CASTPCCRNAGCVDQPDGTEGRCABGEGTLCERNVDCSPDPCHGKCVGSIASFQAC 571
 QY 299 SPGYTGERCODE-----CPVTTYGVIC-----AETCQ 325
 DB 572 APGYTGICSEQVDECRSQPRYGGKCLDVKYLCRCPPTTGVNCEVNIIDDCAASNCT 631
 QY 326 ---CVNGSKCYHVGACLCERAFNGERCEARL-----CPEGLYGIKCDKRCRCP 369
 DB 632 FGVCGRDGINRYD---CVCQPGFTGPLCNVINEINCCASSPCGEGSCVDGGENGFHC--LCP 685
 QY 370 -----CHLENTHS--CHPMSG--ECACKRGMGSLYCNB----- 398
 DB 686 PGLSPPLCLPLANHPHCAHPCSHGVCHDAPGFRVCCEGWSGPRCSOSLADPACBSOPCQ 745
 QY 399 -----TCSPGFYGEACQOI--CS---CONGACDSDVTGK---CTCAPGFRG 436

DB 746 AGGTCTSDGIGFRCTCAAGFGHQCCEVLSPCTPSLCHEGHCEDDPRLATVCSCPFMQGS 805
 QY 437 -----IDSTPOPLGTGYNCS-----ROGCKNDAY-----CSP----- 466
 DB 806 PRCQDVDEGACASPCGPHG--TCTNLPGNFRICHRGYTGPFCDODIDDCPNLHGGS 864
 QY 467 -VDG-----SCYCKXMGHWVDC-----SINCPSTMGFGONTL 497
 DB 865 CQDVGSPSCCLDGFAGPRCARDVDECLSSPCGPGTCTDVASFCAACPPYGGFCEI 924
 QY 498 -----TCQCLNGACNTLID-----TCTCAPGMRGEC----- 525
 DB 925 DLDPQSPSCNNGGTC--VDGVSFSCLCRGYGTGHOYADPCFSRPLCHGICNPTH 982
 QY 526 ---ELPCODGYTGLNCAERCD-----CSHADGCHPTTHGRCCLPMSGVHCD----- 569
 DB 983 PGFECTCREGFTGSCOCNPVDMCSQAPCONGRCVOTGAYCICPBGMSGRICDIQSLPCT 1042
 QY 570 -----SYCAEGRWGPNCSL---PC 585

RESULT 7

ID NTC2_RAT STANDARD; PRT: 2471 AA.
 AC O9QW30;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2).
 GN NOTCH2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_Taxid:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain.
 RX MEDLINE-93202015; PubMed-1295745;
 RA Weinmaster G., Roberts V.J., Lemke G.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RL Development 116:931-941(1992).
 RP [2]
 RP TISSUE SPECIFICITY.
 RX MEDLINE-21331789; PubMed-11438922;
 RA Irvin D.R., Zurcher S.D., Nguyen T., Weinmaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-Delta signaling system during brain
 RL development.";
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May play an essential role in postimplantation
 CC differentiation, probably in some aspect of cell specification and/or
 CC development. (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the spleen and choroid
 CC plexus in the brain. Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within numerous
 CC cells throughout the CNS. It is more highly localized to
 CC ventricular germinal zones. Also found in the heart, liver and
 CC kidney.
 CC -1- DEVELOPMENTAL STAGE: Expressed in the brain during E14 and E17.


```

CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
CC which is proteolytically cleaved by a furin-like convertase in the
CC trypsin-gold network before it reaches the plasma membrane to yield
CC an active, ligand-accessible form. Cleavage results in a C-
CC terminal fragment N(TCM) and a N-terminal fragment N(ETC). Following
CC ligand binding, it is cleaved by TNF-alpha converting enzyme
CC (TACE) to yield a membrane-associated intermediate fragment called
CC notch extracellular truncation (NEXT). This fragment is then
CC cleaved by presenilin dependent gamma-secretase to release a
CC notch-derived peptide containing the intracellular domain (NICD)
CC from the membrane (By similarity).
CC -1- PTM: Phosphorylated (By similarity).
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 35 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M93661; AK13558.1; -
CC HSP: P00743; ICCF.
CC InterPro: IPR002110; ANK.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF-2.
CC InterPro: IPR001881; EGF-Ca.
CC InterPro: IPR001438; EGF-11.
CC InterPro: IPR002049; Laminin_EGF.
CC InterPro: IPR000800; Notch.
CC Pfam: PF00008; EGF: 35.
CC Pfam: PF00023; ank: 6.
CC Pfam: PF00066; notch: 2.
CC PRINTS: PR00010; EGFBLD.
CC PRINTS: PR00011; EGFBLD.
CC PRINTS: PR01452; NOTCH.
CC SMART: SM00248; ANK: 4.
CC SMART: SM00179; EGF-Ca: 22.
CC SMART: SM00001; EGF-like: 10.
CC SMART: SM00084; ANK_REPEAT: 4.
CC PROSITE: PS50088; ANK_REPEAT: 1.
CC PROSITE: PS50297; ANK_REPEAT_REGION: 1.
CC PROSITE: PS00010; ASX_HYDROXYL: 22.
CC PROSITE: PS00022; EGF-1: 34.
CC PROSITE: PS01186; EGF-2: 26.
CC PROSITE: PS01187; EGF-Ca: 22.
CC Receptor, Transcription regulation; Activator; Differentiation;
CC Developmental protein; Repeat; ANK repeat; EGF-like domain;
CC Transmembrane; Glycoprotein; Signal; Phosphorylation.
CC SIGNAL: 1 25
CC CHAIN: 26 2471
CC CHAIN: 1666 2471
CC CHAIN: 1697 2471
CC DOMAIN: 26 1677
CC TRANSMEM: 1678 1698
CC DOMAIN: 1699 2471
CC DOMAIN: 26 63
CC DOMAIN: 64 102
CC DOMAIN: 105 143
CC DOMAIN: 144 180
CC DOMAIN: 182 219
CC DOMAIN: 221 258
CC DOMAIN: 260 296
CC DOMAIN: 298 336
CC DOMAIN: 338 374
CC DOMAIN: 375 413

```

```

FT DOMAIN 415 454 EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 456 492 EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 494 530 EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 532 568 EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 570 605 EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 607 643 EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 645 680 EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 682 718 EGF-LIKE 18, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 720 755 EGF-LIKE 19, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 757 793 EGF-LIKE 20, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 795 831 EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 833 871 EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 873 909 EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 911 947 EGF-LIKE 24, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 949 985 EGF-LIKE 25, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 987 1023 EGF-LIKE 26, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1025 1061 EGF-LIKE 27, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1063 1099 EGF-LIKE 28, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1101 1147 EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1149 1185 EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1187 1223 EGF-LIKE 31, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1225 1262 EGF-LIKE 32, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1264 1302 EGF-LIKE 33, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1304 1342 EGF-LIKE 34, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1344 1412 EGF-LIKE 35, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1414 1448 EGF-LIKE 36, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1450 1484 EGF-LIKE 37, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1486 1520 EGF-LIKE 38, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1522 1556 EGF-LIKE 39, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1558 1592 EGF-LIKE 40, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1594 1628 EGF-LIKE 41, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1630 1664 EGF-LIKE 42, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1666 1700 EGF-LIKE 43, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1702 1736 EGF-LIKE 44, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1738 1772 EGF-LIKE 45, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1774 1808 EGF-LIKE 46, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1810 1844 EGF-LIKE 47, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1846 1880 EGF-LIKE 48, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1882 1916 EGF-LIKE 49, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1918 1952 EGF-LIKE 50, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1954 1988 EGF-LIKE 51, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1990 2024 EGF-LIKE 52, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2026 2060 EGF-LIKE 53, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2062 2096 EGF-LIKE 54, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2098 2132 EGF-LIKE 55, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2134 2168 EGF-LIKE 56, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2170 2204 EGF-LIKE 57, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2206 2240 EGF-LIKE 58, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2242 2276 EGF-LIKE 59, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2278 2312 EGF-LIKE 60, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2314 2348 EGF-LIKE 61, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2350 2384 EGF-LIKE 62, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2386 2420 EGF-LIKE 63, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2422 2456 EGF-LIKE 64, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2458 2492 EGF-LIKE 65, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2494 2528 EGF-LIKE 66, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2530 2564 EGF-LIKE 67, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2566 2600 EGF-LIKE 68, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2602 2636 EGF-LIKE 69, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2638 2672 EGF-LIKE 70, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2674 2708 EGF-LIKE 71, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2710 2744 EGF-LIKE 72, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2746 2780 EGF-LIKE 73, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2782 2816 EGF-LIKE 74, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2818 2852 EGF-LIKE 75, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2854 2888 EGF-LIKE 76, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2890 2924 EGF-LIKE 77, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2926 2960 EGF-LIKE 78, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2962 2996 EGF-LIKE 79, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 2998 3032 EGF-LIKE 80, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3034 3068 EGF-LIKE 81, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3070 3104 EGF-LIKE 82, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3106 3140 EGF-LIKE 83, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3142 3176 EGF-LIKE 84, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3178 3212 EGF-LIKE 85, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3214 3248 EGF-LIKE 86, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3250 3284 EGF-LIKE 87, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3286 3320 EGF-LIKE 88, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3322 3356 EGF-LIKE 89, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3358 3392 EGF-LIKE 90, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3394 3428 EGF-LIKE 91, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3430 3464 EGF-LIKE 92, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3466 3500 EGF-LIKE 93, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3502 3536 EGF-LIKE 94, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3538 3572 EGF-LIKE 95, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3574 3608 EGF-LIKE 96, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3610 3644 EGF-LIKE 97, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3646 3680 EGF-LIKE 98, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3682 3716 EGF-LIKE 99, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3718 3752 EGF-LIKE 100, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3754 3788 EGF-LIKE 101, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3790 3824 EGF-LIKE 102, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3826 3860 EGF-LIKE 103, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3862 3896 EGF-LIKE 104, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3898 3932 EGF-LIKE 105, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3934 3968 EGF-LIKE 106, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3970 4004 EGF-LIKE 107, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4006 4040 EGF-LIKE 108, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4042 4076 EGF-LIKE 109, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4078 4112 EGF-LIKE 110, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4114 4148 EGF-LIKE 111, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4150 4184 EGF-LIKE 112, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4186 4220 EGF-LIKE 113, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4222 4256 EGF-LIKE 114, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4258 4292 EGF-LIKE 115, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4294 4328 EGF-LIKE 116, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4330 4364 EGF-LIKE 117, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4366 4400 EGF-LIKE 118, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4402 4436 EGF-LIKE 119, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4438 4472 EGF-LIKE 120, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4474 4508 EGF-LIKE 121, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4510 4544 EGF-LIKE 122, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4546 4580 EGF-LIKE 123, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4582 4616 EGF-LIKE 124, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4618 4652 EGF-LIKE 125, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4654 4688 EGF-LIKE 126, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4690 4724 EGF-LIKE 127, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4726 4760 EGF-LIKE 128, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4762 4796 EGF-LIKE 129, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4798 4832 EGF-LIKE 130, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4834 4868 EGF-LIKE 131, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4870 4904 EGF-LIKE 132, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4906 4940 EGF-LIKE 133, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4942 4976 EGF-LIKE 134, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 4978 5012 EGF-LIKE 135, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5014 5048 EGF-LIKE 136, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5050 5084 EGF-LIKE 137, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5086 5120 EGF-LIKE 138, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5122 5156 EGF-LIKE 139, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5158 5192 EGF-LIKE 140, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5194 5228 EGF-LIKE 141, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5230 5264 EGF-LIKE 142, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5266 5300 EGF-LIKE 143, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5302 5336 EGF-LIKE 144, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5338 5372 EGF-LIKE 145, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5374 5408 EGF-LIKE 146, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5410 5444 EGF-LIKE 147, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5446 5480 EGF-LIKE 148, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5482 5516 EGF-LIKE 149, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5518 5552 EGF-LIKE 150, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5554 5588 EGF-LIKE 151, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5590 5624 EGF-LIKE 152, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5626 5660 EGF-LIKE 153, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5662 5696 EGF-LIKE 154, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5698 5732 EGF-LIKE 155, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5734 5768 EGF-LIKE 156, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5770 5804 EGF-LIKE 157, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5806 5840 EGF-LIKE 158, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5842 5876 EGF-LIKE 159, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5878 5912 EGF-LIKE 160, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5914 5948 EGF-LIKE 161, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5950 5984 EGF-LIKE 162, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 5986 6020 EGF-LIKE 163, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6022 6056 EGF-LIKE 164, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6058 6092 EGF-LIKE 165, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6094 6128 EGF-LIKE 166, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6130 6164 EGF-LIKE 167, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6166 6200 EGF-LIKE 168, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6202 6236 EGF-LIKE 169, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6238 6272 EGF-LIKE 170, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6274 6308 EGF-LIKE 171, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6310 6344 EGF-LIKE 172, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6346 6380 EGF-LIKE 173, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6382 6416 EGF-LIKE 174, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6418 6452 EGF-LIKE 175, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6454 6488 EGF-LIKE 176, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6490 6524 EGF-LIKE 177, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6526 6560 EGF-LIKE 178, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6562 6596 EGF-LIKE 179, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6598 6632 EGF-LIKE 180, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6634 6668 EGF-LIKE 181, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6670 6704 EGF-LIKE 182, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6706 6740 EGF-LIKE 183, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6742 6776 EGF-LIKE 184, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6778 6812 EGF-LIKE 185, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6814 6848 EGF-LIKE 186, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6850 6884 EGF-LIKE 187, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6886 6920 EGF-LIKE 188, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6922 6956 EGF-LIKE 189, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6958 6992 EGF-LIKE 190, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 6994 7028 EGF-LIKE 191, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7030 7064 EGF-LIKE 192, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7066 7100 EGF-LIKE 193, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7102 7136 EGF-LIKE 194, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7138 7172 EGF-LIKE 195, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7174 7208 EGF-LIKE 196, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7210 7244 EGF-LIKE 197, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7246 7280 EGF-LIKE 198, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7282 7316 EGF-LIKE 199, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7318 7352 EGF-LIKE 200, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7354 7388 EGF-LIKE 201, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7390 7424 EGF-LIKE 202, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7426 7460 EGF-LIKE 203, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7462 7496 EGF-LIKE 204, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7498 7532 EGF-LIKE 205, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7534 7568 EGF-LIKE 206, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7570 7604 EGF-LIKE 207, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7606 7640 EGF-LIKE 208, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7642 7676 EGF-LIKE 209, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7678 7712 EGF-LIKE 210, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7714 7748 EGF-LIKE 211, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7750 7784 EGF-LIKE 212, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7786 7820 EGF-LIKE 213, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7822 7856 EGF-LIKE 214, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7858 7892 EGF-LIKE 215, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7894 7928 EGF-LIKE 216, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7930 7964 EGF-LIKE 217, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 7966 8000 EGF-LIKE 218, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8002 8036 EGF-LIKE 219, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8038 8072 EGF-LIKE 220, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8074 8108 EGF-LIKE 221, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8110 8144 EGF-LIKE 222, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8146 8180 EGF-LIKE 223, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8182 8216 EGF-LIKE 224, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8218 8252 EGF-LIKE 225, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8254 8288 EGF-LIKE 226, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8290 8324 EGF-LIKE 227, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8326 8360 EGF-LIKE 228, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8362 8396 EGF-LIKE 229, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8398 8432 EGF-LIKE 230, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8434 8468 EGF-LIKE 231, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8470 8504 EGF-LIKE 232, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8506 8540 EGF-LIKE 233, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8542 8576 EGF-LIKE 234, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8578 8612 EGF-LIKE 235, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8614 8648 EGF-LIKE 236, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8650 8684 EGF-LIKE 237, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8686 8720 EGF-LIKE 238, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8722 8756 EGF-LIKE 239, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8758 8792 EGF-LIKE 240, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8794 8828 EGF-LIKE 241, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8830 8864 EGF-LIKE 242, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8866 8900 EGF-LIKE 243, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8902 8936 EGF-LIKE 244, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8938 8972 EGF-LIKE 245, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 8974 9008 EGF-LIKE 246, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9010 9044 EGF-LIKE 247, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9046 9080 EGF-LIKE 248, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9082 9116 EGF-LIKE 249, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9118 9152 EGF-LIKE 250, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9154 9188 EGF-LIKE 251, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9190 9224 EGF-LIKE 252, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9226 9260 EGF-LIKE 253, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9262 9296 EGF-LIKE 254, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9298 9332 EGF-LIKE 255, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9334 9368 EGF-LIKE 256, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9370 9404 EGF-LIKE 257, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9406 9440 EGF-LIKE 258, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9442 9476 EGF-LIKE 259, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9478 9512 EGF-LIKE 260, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9514 9548 EGF-LIKE 261, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9550 9584 EGF-LIKE 262, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9586 9620 EGF-LIKE 263, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9622 9656 EGF-LIKE 264, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9658 9692 EGF-LIKE 265, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9694 9728 EGF-LIKE 266, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9730 9764 EGF-LIKE 267, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9766 9800 EGF-LIKE 268, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9802 9836 EGF-LIKE 269, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9838 9872 EGF-LIKE 270, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9874 9908 EGF-LIKE 271, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9910 9944 EGF-LIKE 272, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9946 9980 EGF-LIKE 273, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 9982 10016 EGF-LIKE 274, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10018 10052 EGF-LIKE 275, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10054 10088 EGF-LIKE 276, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10090 10124 EGF-LIKE 277, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10126 10160 EGF-LIKE 278, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10162 10196 EGF-LIKE 279, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10198 10232 EGF-LIKE 280, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10234 10268 EGF-LIKE 281, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10270 10304 EGF-LIKE 282, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10306 10340 EGF-LIKE 283, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10342 10376 EGF-LIKE 284, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10378 10412 EGF-LIKE 285, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10414 10448 EGF-LIKE 286, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10450 10484 EGF-LIKE 287, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10486 10520 EGF-LIKE 288, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10522 10556 EGF-LIKE 289, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10558 10592 EGF-LIKE 290, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10594 10628 EGF-LIKE 291, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10630 10664 EGF-LIKE 292, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10666 10700 EGF-LIKE 293, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10702 10736 EGF-LIKE 294, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10738 10772 EGF-LIKE 295, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10774 10808 EGF-LIKE 296, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10810 10844 EGF-LIKE 297, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10846 10880 EGF-LIKE 298, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10882 10916 EGF-LIKE 299, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10918 10952 EGF-LIKE 300, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10954 10988 EGF-LIKE 301, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 10990 11024 EGF-LIKE 302, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11026 11060 EGF-LIKE 303, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11062 11096 EGF-LIKE 304, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11098 11132 EGF-LIKE 305, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11134 11168 EGF-LIKE 306, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11170 11204 EGF-LIKE 307, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11206 11240 EGF-LIKE 308, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11242 11276 EGF-LIKE 309, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11278 11312 EGF-LIKE 310, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11314 11348 EGF-LIKE 311, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11350 11384 EGF-LIKE 312, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11386 11420 EGF-LIKE 313, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11422 11456 EGF-LIKE 314, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11458 11492 EGF-LIKE 315, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11494 11528 EGF-LIKE 316, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11530 11564 EGF-LIKE 317, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11566 11600 EGF-LIKE 318, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11602 11636 EGF-LIKE 319, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11638 11672 EGF-LIKE 320, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11674 11708 EGF-LIKE 321, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11710 11744 EGF-LIKE 322, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11746 11780 EGF-LIKE 323, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11782 11816 EGF-LIKE 324, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11818 11852 EGF-LIKE 325, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11854 11888 EGF-LIKE 326, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11890 11924 EGF-LIKE 327, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11926 11960 EGF-LIKE 328, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11962 11996 EGF-LIKE 329, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 11998 12032 EGF-LIKE 330, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12034 12068 EGF-LIKE 331, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12070 12104 EGF-LIKE 332, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12106 12140 EGF-LIKE 333, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12142 12176 EGF-LIKE 334, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12178 12212 EGF-LIKE 335, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12214 12248 EGF-LIKE 336, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12250 12284 EGF-LIKE 337, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12286 12320 EGF-LIKE 338, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12322 12356 EGF-LIKE 339, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12358 12392 EGF-LIKE 340, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12394 12428 EGF-LIKE 341, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12430 12464 EGF-LIKE 342, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12466 12500 EGF-LIKE 343, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 12502 12536 EGF-LIKE 344, CAL
```

FT DISULFID 498 509 BY SIMILARITY.
 FT DISULFID 503 518 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 536 547 BY SIMILARITY.
 FT DISULFID 541 556 BY SIMILARITY.
 FT DISULFID 558 567 BY SIMILARITY.
 FT DISULFID 574 584 BY SIMILARITY.

Query Match 19.08; Score 685.5; DB 1; Length 2471;
 Best Local Similarity 24.88; Pred. No. 1.7e-36;
 Matches 218; Conservative 83; Mismatches 240; Indels 339; Gaps 55;

QY 3 ISLNSCLFICL---LLCH-----WIGFASPLNE--DPNVCSHW-----ES 39
 Db 531 IDIDCSSTPLNGAKCIDHNGYECOCATGFTGLCDNDNDPDCPHHGGCQDGDIDS 590
 QY 40 YSVTVOESYPHPF--DOI--YVTS-----CTDLNMFKCT-----RHRSY---- 76
 Db 591 YTCICNPYGMALCSDOIDECYSPLNDGRICIDLNVNGYQCNCPGTSGLNCIEINFDCA 650
 QY 77 RTAVRHGF--KTMVRRKQCCPGFESEEMC--VPHCADR----- 112
 Db 651 SNPLHGACVDJINRYSCVCSGPF--TGQRNIDIDECASNPCKRDATCINDVNGFRMC 708
 QY 113 -----CVHGRG--IAPNTCCBPNGGTFNCSACDGDHMGPHCT 149
 Db 709 PEGRHPSYCQVNECLSPCIHNGTCTGLSGYKCLCAGVNGINCE--VDKN-----ECL 762
 QY 150 SRCQCKNGLCNPTGA--CHCAAGFRGWCE--DRC-----EQTYYGNDCH--QRCQC 197
 Db 763 SN-PONGNGTNNLVNGYRCKTKGFKGYNCQVNIDECASNPCLNOGTCLDVSGYCHC 821
 QY 198 Q---NGATCDHYTGRCRPPRYTGAFCED-----LCPRGKHGPGCE---QRC---P 239
 Db 822 MLPTGKACQIVYLAFCSPNCEANAYVKEAPNFESFTCLCAPGOWGQCTYVDVDECVSKP 881
 QY 240 QONGVCCHHVYGE--CSCPSSGMGTVCQPCPEGRFGRKNCSEOC--OCHNGGTC--DAA 292
 Db 882 CMNNGICHTNGQSYWCCECPGPGSGMDCEEDI-----NDCLANCGNGSGVDKVN 931
 QY 293 TGGCHSGYTGGERCODE-----CVGYTYGVC---AETC 324
 Db 932 TFCSCILPFGVGDCKQDTMNECLSEPCKNNGTCSDYVNSTCTCPAGHGVHCENNIDEC 991
 QY 325 ---OCVNGKCYHVS---ACICEAGFAGERC-----EARLPEGLYGIKC 364
 Db 992 TESSCFNGGTC--VDGINSFSCILCFVGTGTPCLHDINECSSNPCLNSGTCDGLGTYRC 1049
 QY 365 -----DKRC-----PCHLENTSHCHPMSGECACRPGMSGLYCNE----- 398
 Db 1050 TCPLEGYTKNCOITLVNLCSPBCKNKGTCAQAEKARPRCLCPGMDGAGAVDLVNSCRMAA 1109
 QY 399 -----TCSPPGYEACQO---TCS---QNGACDSTVYG 426
 Db 1110 LQKGVPEHLCOHSGICINAGNTHHCQPLGTGSGYCEQDEDCASNPCCOHGATCSDFTIG 1169
 QY 427 --KCTCAFGKFGIDCSTPCPLGTGINSRCCNDNAVCSPYDOSCTCKAKMGHVDSI 484
 Db 1170 GYRCGVGTGVGVNCE-----YVDECOQNPCCNGGTCTIDLNVHFKCS----- 1212
 QY 485 RCPSTGWFQGC--NL--TC-----OCLNGAGC--NTLDG--TCTCAPGWRGKECE-----L 527
 Db 1213 -CPETRBLLEENIDDAAGAPHCINGGQCYDRIGYSCRCLPGRAGRCEDINECLSN 1271
 QY 528 PC-QDGYGLNCAE-----KDCSHA-----DGCH-----PTTG 555
 Db 1272 PCSSGGS--LDCIQLKNNYQCVCSRSAFGRHCEFLDVCPOKPCLINGTCVAVNSPDGF 1329
 QY 556 HCRCLPGMSVGHVCSVCAEGRM-----GRNCSLP 584
 Db 1330 ICRCPGFGSGARCOSSCGOVKRCRGEQCVHAGSGPHCFP 1369

RESULT 8
 NTG3_RAT
 ID NTG3_RAT STANDARD; PRT; 2319 AA.
 AC 09R172;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 3 precursor (Notch 3).
 GN NOTCH3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Haritunians T., Boulter J., Weimaster G., Schanen N.C.;
 RT "Rattus norvegicus mRNA for Notch 3."
 RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP FUNCTION.
 RA MEDLINE-21094508; PubMed-11182080;
 RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kurooka H.,
 RA Honjo T.;
 RT "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent
 RT neural progenitor cells to an astroglial fate."
 RL Neuron 29:45-55(2001).
 RN [3]
 RP TISSUE SPECIFICITY.
 RA MEDLINE-21331789; PubMed-11438922;
 RA Iryin D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
 RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple
 RT functional roles for the Notch-DSL signaling system during brain
 RT development."
 RL J. Comp. Neurol. 436:167-181(2001).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (By similarity). Acts instructively to control
 CC the cell fate determination of CNS multipotent progenitor cells,
 CC resulting in astroglial induction and neuron/oligodendrocyte
 CC suppression.
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- TISSUE SPECIFICITY: Expressed in postnatal central nervous system
 CC (CNS) germinal zones and, in early postnatal life, within
 CC numerous cells throughout the CNS. It is more highly localized
 CC to ventricular germinal zones.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEX). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

Query Match	Best Local Similarity	Matches 219; Conservative	Score 682.5; DB 1; Length 2319;	Indels 415; Gaps 49;
94	CCPFEYEGSGMCVPHCADKCVHGRIAPNT-----CCCEPROMGTGNCSSACDGDHW	144		
226	CLPGEF--EGQNCSEVN-YDDCEPHRCILNGSTGVGDVNTNYCCGPPEMTGOFCTEDVD-----	278		
145	GPHTCSRCQ-----CKNGALCNPTTG--ACHCAAG-----FGMRCE	179		
279	-----ECOLPQNMACHNGTCFNLIGGHSVCVNGMTGESQSNIDCATAVCFHGATCH	332		
180	DR-----CGQGYTGANDCH--QRC--QCQNGATCD--HTTGE--CPCPGCYGAFED	223		
333	DRVASFYCACPMKGTGLCLCHLDACVSNPCHEDALCIDNPVSGRAICTCPPEFTGGACDQ	392		

```

OY 224 -----LCPGKHGPOCE----- 235
DB 333 DVDECSIGANPCENHIGRCVNTGSGFLCQGRGRTYTPRCCTDVNECLSGPCRNQATCLDRI 452
OY 236 -----QRCPCQNGVC--HHVTG--ECSGPGSMGTVC----- 264
DB 453 GQFTICMAGFTGTCFCEVDIDECQSSPCVNGVCCKDRVNGFECTCPSGSSGTCQLDVE 512
OY 265 -----GQP-----CPRGRKNCQSQ---EQQ---CHNGTCTDA--ATGQCHC 298
DB 513 CASTPCRNAGKACVDPDQDYECRCAGFEFTLCERNVDDCSPPCRHGRGVDDIASFSKAC 572
OY 299 SPGTGGERCODE-----CPVGRGYVIC-----AETCQ 325
DB 573 ABGTGIRCESQVDECRSPRCRYGGKCLDLYDKTLCRCPPGTGVNCEYINDCCASNPT 632
OY 336 ---CVNGRCXYHVSAGACLCCEAGFAGEKCEARL-----CPEGLYKICKDKRCP 369
DB 633 FCVCBDGINRYD-----CYCQPGFTGPLCINVEINECASSPCGSGGSCYDGENGFHC--LCP 686
OY 370 -----CHLENTHS--CHPMSSG--ECACKRFGWSGLYCNF----- 398
DB 687 PGLSLPCLIPANHPCAHNRKPSHGVCCHDAPGFGQCVCDPMWSGPRCSQSLAPDACESOPQ 746
OY 399 -----TCSRFYGEACQOI--CS---CONGADDSVTGK---CTCAPRFGK 436
DB 747 AGGTCTSDGIGHCACAFEGQHOCEVLSFCTPSLCERHGHCESDPDQLTYVSCFPGWOG 806
OY 437 IDC-----STPC-PLGT--YGINCSRCGCKNDAY-----CSP----- 466
DB 807 PRCQDQVDECAAGSPCGPHGTCTNLPSPRCICHGTYGTFPCDQDIDDDPRLCNGSGC 866
OY 467 VDG-----SCTCAAGHGVDC-----SIRCSGSMWAGCN-- 496
DB 867 QGVGSFSCSLSGFAGPCARDVDECLSSPCGPGCTDHDVASFCTCPCPGYGFHCETD 926
OY 497 -LTC---OCLNGACNTLDG---TCTCAPGWRGEC----- 525
DB 927 LLDGSPSSCFNGCTC--VDGVNSFSCICRFGYTGTHCQYKVPDPCRSRPLHGCINPHTS 984
OY 536 -ELPCQDGTGVLNCAERCD-----CSHADGCHPTTGHCRLPMSGVHCD----- 569
DB 985 GFECTCRGEGTNGOCNPNVDMCSQAPCONGRCVOTGATVACICRPMWSGLDIPSLPCTE 1044
OY 570 -----SYCAEGRNGPNCSL---PC 585
DB 1045 AAAMGVRLQLCQAGGOCIDKDSHSHYCVCPGRMGSHCEDEVDPC 1090

```

RESULT 9

```

ID TENX_HUMAN STANDARD: PRT: 4289 AA.
AC P22105; P78530; P78531; Q08424; Q90M67;
DT 01-AUG-1991 (Rel. 19, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tenascin-X precursor (TN-X) (hexabrachion-like).
GN TNXB OR TNX OR XB OR HMBL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Schwartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus."
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-747 AND 1687-1944 FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=93300909; PubMed=7686164;

```

```

RA Bristow J., Tee M.K., Gitelman S.E., Mellon S.H., Miller W.L.;
RT "Tenascin-X: a novel extracellular matrix protein encoded by the human
RT XB gene overlapping P450C21B."
RL J. Cell Biol. 122:265-278(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM XB-SHORT).
RC TISSUE=Adrenal gland;
RX MEDLINE=96015044; PubMed=8530023;
RA Tee M.K., Thomson A.A., Bristow J., Miller W.L.;
RT "Sequences promoting the transcription of the human XA gene
RT overlapping P450C21A correctly predict the presence of a novel,
RT adrenal-specific, truncated form of tenascin-X."
RL Genomics 28:171-178(1995).
RN [4]
RP SEQUENCE OF 1-23 FROM N.A.
RC TISSUE=Fetal adrenal gland;
RX MEDLINE=97081760; PubMed=8923003;
RA Speck M., Barry F., Miller W.L.;
RT "Alternate promoters and alternate splicing of human tenascin-X, a
RT gene with 5' and 3' ends buried in other genes."
RL Hum. Mol. Genet. 5:1749-1758(1996).
RN [5]
RP SEQUENCE OF 3470-4289 FROM N.A.
RX MEDLINE=89367293; PubMed=2475872;
RA Morel Y., Bristow J., Gitelman S.E., Miller W.L.;
RT "Transcript encoded on the opposite strand of the human steroid 21-
RT hydroxylase/complement component C4 gene locus."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6582-6586(1989).
RN [6]
RP DISEASE.
RA Published-11642233;
RA Schallert J., Zweers M.C., Steijlen P.M., Dean W.B., Taylor G.,
RA van Vlijmen I.M., van Haren B., Miller W.L., Bristow J.;
RT "A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X
RT deficiency."
RL New Engl. J. Med. 345:1167-1175(2001).
CC -1- FUNCTION: APPEARS TO MEDIATE INTERACTIONS BETWEEN CELLS AND THE
CC EXTRACELLULAR MATRIX. SUBSTRATE-ADHESION MOLECULE THAT APPEARS TO
CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC OF EPITHELIAL TUMORS.
CC -1- SUBCELLULAR LOCATION: Extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; XB (SHOWN HERE) AND
CC XB-SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN FETAL ADRENAL, IN FETAL
CC TESTIS, FETAL SMOOTH, STRIATED AND CARDIAC MUSCLE. XB-SHORT IS
CC ONLY EXPRESSED IN THE ADRENAL GLAND.
CC -1- DISEASE: ASSOCIATION WITH CONGENITAL ADRENAL HYPERPLASIA.
CC -1- DISEASE: Defects in TNXB are the cause of an autosomal recessive
CC form of Ehlers-Danlos syndrome (EDS) manifesting with
CC hyperextensible skin, hypermobile joints, and tissue fragility as
CC observed in classical EDS, but lacking atrophic scars and delayed
CC wound healing.
CC -1- SIMILARITY: CONTAINS 19 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 32 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 FIBRINOGEN C-TERMINAL DOMAIN.
CC -1- CAUTION: THERE ARE TWO GENES FOR TN-X: TNXA AND TNXB. TNXA IS A
CC PARTIAL GENE WHICH CAN SOMETIMES RECOMBINE WITH TNXB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U89337; AAB47488.1; -
DR EMBL: AF019413; AAB67981.1; -
DR EMBL: X71923; CA50739.1; -
DR EMBL: Y13782; CA474109.1; -
DR EMBL: Y13783; CA474110.1; -
DR EMBL: U24488; AAB41287.1; -
DR EMBL: U52696; AAC50889.1; -

```

DR EMBL; M25813; AAA5884.1; -
 DR HSSP; P02671; 1F2D.
 DR Genew; HGNC; 11976; TNXB.
 DR MIM; 600985; -
 DR MIM; 606408; -
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00008; EGF; 15.
 DR Pfam; PF00041; fn3; 33.
 DR Pfam; PF00147; fibrinogen_C; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00181; EGF; 8.
 DR SMART; SM00001; EGF_Like; 1.
 DR SMART; SM00186; FBC; 1.
 DR SMART; SM00060; FN3; 31.
 DR PROSITE; PS00022; EGF_1; 18.
 DR PROSITE; PS01186; EGF_2; 19.
 DR PROSITE; PS00514; FIBRIN_AG_C DOMAIN; 1.
 DR Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Colled coll;
 KW Extracellular matrix; Alternative splicing; Signal;
 KW Ehlers-Danlos syndrome;
 FT SIGNAL 1 23
 FT CHAIN 24 4289
 FT DOMAIN 156 168 EGF-Like 1 (INCOMPLETE).
 FT DOMAIN 183 213 EGF-Like 2.
 FT DOMAIN 214 244 EGF-Like 3.
 FT DOMAIN 245 275 EGF-Like 4.
 FT DOMAIN 276 306 EGF-Like 5.
 FT DOMAIN 307 337 EGF-Like 6.
 FT DOMAIN 338 368 EGF-Like 7.
 FT DOMAIN 369 399 EGF-Like 8.
 FT DOMAIN 400 430 EGF-Like 9.
 FT DOMAIN 431 461 EGF-Like 10.
 FT DOMAIN 462 492 EGF-Like 11.
 FT DOMAIN 493 523 EGF-Like 12.
 FT DOMAIN 524 554 EGF-Like 13.
 FT DOMAIN 555 585 EGF-Like 14.
 FT DOMAIN 586 616 EGF-Like 15.
 FT DOMAIN 617 647 EGF-Like 16.
 FT DOMAIN 648 679 EGF-Like 17.
 FT DOMAIN 684 714 EGF-Like 18.
 FT DOMAIN 715 746 EGF-Like 19.
 FT DOMAIN 792 872 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 901 922 COLLED COLL.
 FT DOMAIN 941 1021 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 1047 1127 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 1149 1226 FIBRONECTIN TYPE-III 4.
 FT DOMAIN 1246 1327 FIBRONECTIN TYPE-III 5.
 FT DOMAIN 1348 1429 FIBRONECTIN TYPE-III 6.
 FT DOMAIN 1459 1540 FIBRONECTIN TYPE-III 7.
 FT DOMAIN 1561 1642 FIBRONECTIN TYPE-III 8.
 FT DOMAIN 1659 1736 FIBRONECTIN TYPE-III 9.
 FT DOMAIN 1736 1836 FIBRONECTIN TYPE-III 10.
 FT DOMAIN 1836 1939 FIBRONECTIN TYPE-III 11.
 FT DOMAIN 1962 2039 FIBRONECTIN TYPE-III 12.
 FT DOMAIN 2069 2150 FIBRONECTIN TYPE-III 13.
 FT DOMAIN 2167 2248 FIBRONECTIN TYPE-III 14.
 FT DOMAIN 2266 2347 FIBRONECTIN TYPE-III 15.
 FT DOMAIN 2471 2552 FIBRONECTIN TYPE-III 16.
 FT DOMAIN 2552 2633 FIBRONECTIN TYPE-III 17.
 FT DOMAIN 2688 2769 FIBRONECTIN TYPE-III 18.
 FT DOMAIN 2769 2850 FIBRONECTIN TYPE-III 19.
 FT DOMAIN 2889 2972 FIBRONECTIN TYPE-III 20.
 FT DOMAIN 2997 3078 FIBRONECTIN TYPE-III 21.
 FT DOMAIN 3078 3166 FIBRONECTIN TYPE-III 22.
 FT DOMAIN 3211 3292 FIBRONECTIN TYPE-III 23.
 FT DOMAIN 3307 3384 FIBRONECTIN TYPE-III 24.
 FT DOMAIN 3384 3481 FIBRONECTIN TYPE-III 25.
 FT DOMAIN 3481 3575 FIBRONECTIN TYPE-III 26.
 FT DOMAIN 3575 3682 FIBRONECTIN TYPE-III 27.
 FT DOMAIN 3601 3682 FIBRONECTIN TYPE-III 28.

FT DOMAIN 3699 3787 FIBRONECTIN TYPE-III 29.
 FT DOMAIN 3801 3879 FIBRONECTIN TYPE-III 30.
 FT DOMAIN 3890 3971 FIBRONECTIN TYPE-III 31.
 FT DOMAIN 3978 4059 FIBRONECTIN TYPE-III 32.
 FT DOMAIN 4071 4289 FIBRONECTIN TYPE-III 33.
 FT SITE 1748 1750 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 187 197 BY SIMILARITY.
 FT DISULFID 191 202 BY SIMILARITY.
 FT DISULFID 204 213 BY SIMILARITY.
 FT DISULFID 218 228 BY SIMILARITY.
 FT DISULFID 222 233 BY SIMILARITY.
 FT DISULFID 235 244 BY SIMILARITY.
 FT DISULFID 249 259 BY SIMILARITY.
 FT DISULFID 253 264 BY SIMILARITY.
 FT DISULFID 266 275 BY SIMILARITY.
 FT DISULFID 280 290 BY SIMILARITY.
 FT DISULFID 284 295 BY SIMILARITY.
 FT DISULFID 297 306 BY SIMILARITY.
 FT DISULFID 311 321 BY SIMILARITY.
 FT DISULFID 315 326 BY SIMILARITY.
 FT DISULFID 328 337 BY SIMILARITY.
 FT DISULFID 342 352 BY SIMILARITY.
 FT DISULFID 346 357 BY SIMILARITY.
 FT DISULFID 359 368 BY SIMILARITY.
 FT DISULFID 373 383 BY SIMILARITY.
 FT DISULFID 377 388 BY SIMILARITY.
 FT DISULFID 390 399 BY SIMILARITY.
 FT DISULFID 404 414 BY SIMILARITY.
 FT DISULFID 408 419 BY SIMILARITY.
 FT DISULFID 421 430 BY SIMILARITY.
 FT DISULFID 435 445 BY SIMILARITY.
 FT DISULFID 439 450 BY SIMILARITY.

Query Match 18.9%; Score 682; DB 1; Length 4289;
 Best Local Similarity 28.4%; Pred. No. 4,28-36;
 Matches 191; Conservative 41; Mismatches 178; Indels 262; Gaps 38;

QY 94 CCEPFYISG-----ENCYPHCKADKCVH-----RCIAPNTCCOEPGWSGTCSSACDG 141
 DB 125 CCPASAGAGTQDVTVRLC-----SLHGVFDLSRC-----TSCCEPWSGSGTSDPTDA 173
 QY 142 D-----HMGPHC-----TSROCKKAGALCNPTYG 165
 DB 174 EIPSSPPSASGSCPPDDCNDGRCVRCVRCPEPTGSCGWPSCPPDCQGRRC--VQG 231
 QY 166 ACHCAAGFRGWRCEDR-CEGTYGNDCHORCOGATCDHYTGECRCPPYTGAFD-ED 223
 DB 232 VCVCRAAGFSQDSCSQRSCPRG-----CSQGRRCG-----GRVCDDPRTYTDGDMR 278
 QY 224 LCPRGKRGPOCEQRCPQNGVGCHHTVGECSGPGMMGYVCG-OPCEBGRGKNCQEQ 282
 DB 279 SCPRG-----CSQGRGEN-----GRVCNPGYGEEDCGVSCPRG-----CSQGR 320
 QY 283 CHNGGTADTAAGOCSPGTYGERC-ODECPVGYGVLCATETOCVNGKCYHVSACLC 341
 DB 321 CKD-----GRVCDDPRTYGEEDCGTRSCP-----WDGSGGRG--VDGRVCV 359
 QY 342 EAGPAGERCEARLCPD-----GLYGIKCDKR-CPCHLENTHSCHPMSG 383
 DB 360 WPGYTGEDCSTRPCRCRGRGRCDEDECICDGYSSDDGVSHPDCDQGRRC--DG 417
 QY 384 ECACKPQMSLYCNE-----TCSGPFYGERC-QQIC--SCQNGADCD 422
 DB 418 RCVCMPGYTGTDCGSRACPDRCRGRGRCENGVCVNAVSGEDCGVNSCGDGRGRCE 477
 QY 423 SVTKGTCAGFGKGIQST-----PCPLGYIGINCSS-RC--GGKND 461
 DB 478 S--GRCKMCPGYTGRDCTGACPDRCRGRGRCVYDGRVCNPGFTGEDCGSRRCPPGDCRGH 535
 QY 462 AVCSPYDGSCTCRAGHGVDCSIR-CPSGTGWGFCNLTCQCLNG----- 504
 DB 536 GLCE--DGYCVYDAGISGEDCSTRSCPGGCRGNG-----QCLDRCVCEJGYSGEDCGVR 568

```

OY 505 -----GACNTLDGTCTCAPGWRGKCELP-----CODGTGTLN 537
DB 589 OCPNDCSQHGYCQ--DGVICICMEGYVEDCSIRCPNCHGRGCEBRCLODPYTGPT 646
OY 538 CAER---CDCSHADGCHPTTGHCRCLPGWGVHC-----DSV 571
DB 647 CATRCCPADCCRRGRRC--VQGVCLCHVGYGEGDCQEBPPASACPGCGCPRELCRAGOCV 704
OY 572 CAEGWGNCSL 583
DB 705 CVEGRGPDCAI 716

RESULT 10
NTC2_MOUSE
ID NTC2_MOUSE STANDARD: PRT: 2470 AA.
AC 035516; 080941; 006008;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
  B).
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
  single copy of mouse Notch2 gene.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RX MEDLINE-93178563; PubMed-8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
  wide variety of tissues.";
RL Exp. Cell Res. 204:364-372(1993).
RN [3]
RP SEQUENCE OF 1765-2153 FROM N.A.
RX MEDLINE-97075110; PubMed-8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
  Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
RN [4]
RP FUNCTION.
RX MEDLINE-99396706; PubMed-10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
  Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
  embryonic lethality.";
RL Development 126:3415-3424(1999).
RN [5]
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RX MEDLINE-9533893; PubMed-7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
  mouse brain.";
RL Brain Res. Mol. Brain Res. 29:263-272(1995).
RN [6]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (NL-4) undergo presenilin-dependent
  proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

```

```

RX MEDLINE-21374376; PubMed-11459941;
RA Mutant T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
  among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
  Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
  Upon ligand activation through the released notch intracellular
  domain (NICD) it forms a transcriptional activator complex with
  RBP-J kappa and activates genes of the enhancer of split locus.
  Affects the implementation of differentiation, proliferation and
  apoptotic programs (By similarity). May play an essential role in
  postimplantation development, probably in some aspect of cell
  specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
  terminal fragment N(EC) which are probably linked by disulfide
  bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
  proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms: 1 (shown here) and 2; may be
  produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
  neuroepithelia, somites, optic vesicles and branchial arches, but
  not heart.
CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
  the postnatal ependymal cells, and the choroid plexus throughout
  embryonic and postnatal development.
CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
  which is proteolytically cleaved by a furin-like convertase in the
  trans-Golgi network before it reaches the plasma membrane to yield
  an active, ligand-accessible form. Cleavage results in a C-
  terminal fragment N(TW) and a N-terminal fragment N(EC). Following
  ligand binding, it is cleaved by TNF-alpha converting enzyme
  (TACE) to yield a membrane-associated intermediate fragment called
  notch extracellular truncation (NEXT). This fragment is then
  cleaved by presenilin dependent gamma-secretase to release a
  notch-derived peptide containing the intracellular domain (NICD)
  from the membrane.
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D32210; BAA22094.1; -
DR EMBL: X68279; CAA48340.1; -
DR EMBL: U31881; AAC52924.1; -
DR HSSP: P16109; IFSB.
DR MGD: MGI:97364; Notch2.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-1like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF-IT.
DR InterPro: IPR000800; Notch.
DR Pfam: PF00008; EGF; 35.
DR Pfam: PF00023; ank; 6.
DR Pfam: PF00066; notch; 2.
DR PRINTS: PRO0010; EGFBLD0.
DR PRINTS: PRO1452; NOTCH.
DR SMART: SM00248; ANK; 4.
DR SMART: SM00179; EGF_CA; 22.
DR SMART: SM00001; EGF_1like; 12.
DR SMART: SM00004; NL; 3.

```


DR PROSITE: PS50088; ANK_REPEAT: 4.
 DR PROSITE: PS50297; ANK_REPEAT: 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 22.
 DR PROSITE: PS00022; EGF_1; 33.
 DR PROSITE: PS01186; EGF_2; 27.
 DR PROSITE: PS01187; EGF_CA; 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 FT SIGNAL 1 25
 FT CHAIN 26 2470
 FT CHAIN 1666 2470
 FT CHAIN 1697 2470
 FT CHAIN 26 1677
 FT TRANSMEM 1678 1698
 FT DOMAIN 1699 2470
 FT DOMAIN 26 63
 FT DOMAIN 64 102
 FT DOMAIN 105 143
 FT DOMAIN 144 180
 FT DOMAIN 182 219
 FT DOMAIN 221 256
 FT DOMAIN 258 294
 FT DOMAIN 296 334
 FT DOMAIN 336 372
 FT DOMAIN 373 411
 FT DOMAIN 413 452
 FT DOMAIN 454 490
 FT DOMAIN 492 528
 FT DOMAIN 530 566
 FT DOMAIN 568 603
 FT DOMAIN 605 641
 FT DOMAIN 643 678
 FT DOMAIN 680 716
 FT DOMAIN 718 753
 FT DOMAIN 755 791
 FT DOMAIN 829 869
 FT DOMAIN 871 907
 FT DOMAIN 909 945
 FT DOMAIN 947 983
 FT DOMAIN 985 1021
 FT DOMAIN 1023 1059
 FT DOMAIN 1061 1097
 FT DOMAIN 1099 1145
 FT DOMAIN 1147 1183
 FT DOMAIN 1185 1221
 FT DOMAIN 1223 1260
 FT DOMAIN 1262 1300
 FT DOMAIN 1302 1345
 FT DOMAIN 1372 1410
 FT REPEAT 1418 1454
 FT REPEAT 1501 1533
 FT REPEAT 1501 1533
 FT REPEAT 1825 1869
 FT REPEAT 1874 1903
 FT REPEAT 1907 1937
 FT REPEAT 1941 1970
 FT REPEAT 1974 2003
 FT REPEAT 2007 2036
 FT DOMAIN 1645 1648
 FT DOMAIN 1992 1995
 FT DOMAIN 2183 2189
 FT DOMAIN 2183 2189
 FT DOMAIN 2425 2428
 FT DOMAIN 2445 2450
 FT DISULFID 28 41
 FT DISULFID 35 51
 FT DISULFID 53 62
 FT DISULFID 68 79
 Query Match 18.8%; Score 677; DB 1; Length 2470;
 Best Local Similarity 24.7%; Pred. No. 5.8e-36;
 Matches 220; Conservative 78; Mismatches 246; Indels 348; Gaps 56;

QY 3 ISLNSCLSFICL-----LLCH-----WIGTASPLNIE--DPNVCSHW-----ES 39
 Db 529 IDIDCSSTPLNGAKKCIDHPNGYECQCATGTTGLDCEINIDNCDPDCCHGGQODGIDS 588
 QY 40 YSVAYQSEYPRHP--DQI--YITS-----CTDILNFKCT-----RHRVSY----- 76
 Db 589 YTCICNPYMAICSDQIDECYSSPCLNDGRICIDLVNQCNCQPGTGLNCEINFDCA 648
 QY 77 RTAYRHG--EKMTYRRKSOCCPGF-----YESGEMCV----- 106
 Db 649 SNPCRHGVCYGINRYSCVSPFGTGQCNIDIDECASNPRKATCTINDVNGFRCTIPE 708
 QY 107 ----PHC-----ADKCVHGR--TAPTCQEPGSGTNCSSACDGRHWPCHTSR 151
 Db 709 GPHHPSCTSYQNECLSNPCINHGNTGGLSKCLCADAGVGNCE--YDKN----ECLSN 762
 QY 152 CQCKNGLCNITGA--CHCAAGFRMGCE--DRC-----EGTYGNDCH--QRCCQ-- 198
 Db 763 -PCQNGGTNNLVNGYRCTCKRGKYNQVNIDECASNPLNLTGCTEDDVSQYTCHEML 821
 QY 199 --NGATCDHYGECRCPPGYGAFCE-----LCPPGKHGPOCE--QRC--PCQ 241
 Db 822 PYTKNQFVLAPSPNCEAAYCKEAPNRESFCLCAPGMOGKRCTVDVDECISKPCM 881
 QY 242 NGVYCHVTGE--CSPSGMNGVCGQPCPGRGKNCQDRC--QCHNGTC--DATG 294
 Db 882 NNGVCHVTGSGYVCECPGFGMCEEDI-----NCLANPCQNGSCVDHYNTF 931
 QY 295 QCHSPYTGRCODE-----CPRTGYGLC--AEHC-- 324
 Db 932 SCQHPFGIGKCTQTDNNECLSEPCCKNGTCSYVNSYTCCTPAGFGVHCENNIDCTE 991
 QY 325 -QCVNGSKCYHVS--ACICEAGFAG-----ERCEAR-----LC 355
 Db 992 SSCENGGTC--YDINSFSCLPVGTGPFCLHDINCESSNPLCNACTGVDGLTYKIC 1049
 QY 356 PEGTIGKCD-----KRCQCHLEHTSCHPMSECAKPGWSGLYNE----- 398
 Db 1050 PLGTGKNCQTLVNLCSRSPCKNKGTCVQEKARPHCLCPGMDAYCDVLVNSCKAALQ 1109
 QY 399 -----TSGPFGFGEACQO--ICS--CONGADDSVYG-- 426
 Db 1110 KGVVHEHLCHQSHGICINAGNTHHCQCPVLTGTTGTCSEQDLECCASNPCQHGHTCDEFTG 1169
 QY 427 KCTCAPGFKIDSTPCPLGTYGINSRCGCKNDAYCSPVDGSCCTKAGMHGVDCSIRC 486
 Db 1170 RCECVPGYGVNCE-----YEVDCEQNPCCQNGGTCIDLVNHFKS-----C 1211
 QY 487 PSGTWGFGC--NL--TC-----QCLNGAC--NTLDG--TCTCAPGMRGECE-----LPC 529
 Db 1212 PPGTRGLICEINIDECAGPHCLNGGQCVDRIGGYTRCLPFGAGECEDINECLSNPC 1271
 QY 530 -ODGTGLNCAE-----RCDCSHA-----DGC-----HPTTGHG 557
 Db 1272 SSBS--LDCVQLKNNYNCICRSALFTGRHCETFLDVCPPKRLNGTCAVAANMPGFIG 1329
 QY 558 RCLPGMSGVHCDVCAAGRW-----GPNC-----SLPC 585
 Db 1330 RCPGFSGARLQSSCGGVKCRREGQCIHMDSGRCICLANKDCDESGCASNPC 1381
 RESULT 11
 NOTC_DROME STANDARD; PRT: 2703 AA.
 AC P07207: P04154; 097458; 09M4T8;
 DT 01-NOV-1986 (rel. 03, created)
 DT 15-JUN-2002 (rel. 41, last sequence update)
 DT 15-JUN-2002 (rel. 41, last annotation update)
 DE Neurogenic locus Notch protein precursor
 GN N OR EG:140G11.1 OR EG:163A10.2 OR CG3936.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

RX MEDLINE:99180765; PubMed:10079256;
 RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-L.,
 RA Banks A., Leiman J., Ward D., Ish-Horowitz D., Arcavanis-Tsakonas S.,
 RT "Human ligands of the Notch receptor.";
 RL Am. J. Pathol. 154:785-794(1999).
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged1, Jagged2 and Delta1 to regulate cell-fate determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs. May regulate branching morphogenesis in the
 CC developing vascular system (By similarity).
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
 CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1 (shown here), 2 and 3; may be
 CC produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Highly expressed in the heart, moderately in
 CC the lung and placenta and at low levels in the liver, skeletal
 CC muscle, kidney, pancreas, spleen, lymph node, thymus, bone marrow
 CC and fetal liver. No expression was seen in adult brain or
 CC peripheral blood leukocytes.
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TW) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane (By similarity).
 CC -1- PTM: Phosphorylated (By similarity).
 CC -1- POLYMORPHISM: The poly-leu region of NOTCH4 (in the signal
 CC peptide) is polymorphic and the number of Leu varies in the
 CC population (from 6 to 12).
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC -1- SIMILARITY: CONTAINS 28 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 CC in position 1438 to 1463.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D63395; BAA09708.1; ALT_FRAME.
 DR EMBL: D86566; BA131116.1; -
 DR EMBL: U95299; AAC32288.1; -
 DR EMBL: U89335; AAC63097.1; -
 DR EMBL: AB023961; BAB20317.1; -
 DR EMBL: AB024520; BAA88951.1; -
 DR EMBL: AB024578; BAA88952.1; -
 DR HSSP: P08709; 1BP9
 DR Genew: HGNC:7884; NOTCH4.
 DR MIM: 164951; -
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001861; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR000800; Notch.
 DR Pfam: PF00008; EGF; 26.

DR Pfam: PF00023; ank; 6.
 DR Pfam: PF00066; notch; 2.
 DR PRINTS: PR00010; EGFBLD.
 DR PRINTS: PR00011; EGFAMININ.
 DR PRINTS: PR00012; FNTYPEI.
 DR SMART: SM00248; ANK; 5.
 DR SMART: SM00179; EGF_Ca; 11.
 DR SMART: SM00001; EGF_Like; 15.
 DR SMART: SM00004; NL; 2.
 DR PROSITE: PS50088; ANK_REPEAT; 5.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 11.
 DR PROSITE: PS00022; EGF_1; 28.
 DR PROSITE: PS01186; EGF_2; 21.
 DR PROSITE: PS01187; EGF_Ca; 9.
 KW Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Polymorphism;
 KW Triplet repeat expansion; Alternative splicing.
 FT SIGNAL 1 23
 FT CHAIN 24 2003
 FT CHAIN 1432 2003
 FT CHAIN 1467 2003
 FT DOMAIN 24 1447
 FT TRANSMEM 1448 1468
 FT DOMAIN 1469 2003
 FT DOMAIN 24 63
 FT DOMAIN 64 115
 FT DOMAIN 118 155
 FT DOMAIN 156 192
 FT DOMAIN 194 232
 FT DOMAIN 234 272
 FT DOMAIN 276 312
 FT DOMAIN 314 353
 FT DOMAIN 355 391
 FT DOMAIN 392 430
 FT DOMAIN 432 473
 FT DOMAIN 475 511
 FT DOMAIN 513 549
 FT DOMAIN 551 587
 FT DOMAIN 589 625
 FT DOMAIN 626 659
 FT DOMAIN 661 689
 FT DOMAIN 691 727
 FT DOMAIN 729 765
 FT DOMAIN 767 803
 FT DOMAIN 806 842
 FT DOMAIN 844 880
 FT DOMAIN 882 928
 FT DOMAIN 930 966
 FT DOMAIN 968 1004
 FT DOMAIN 1006 1044
 FT DOMAIN 1046 1085
 FT DOMAIN 1087 1126
 FT DOMAIN 1130 1171
 FT DOMAIN 1172 1212
 FT REPEAT 1213 1246
 FT REPEAT 1247 1286
 FT REPEAT 1287 1328
 FT REPEAT 1329 1365
 FT REPEAT 1366 1398
 FT REPEAT 1399 1432
 FT REPEAT 1433 1467
 FT REPEAT 1468 1502
 FT REPEAT 1503 1537
 FT REPEAT 1538 1572
 FT REPEAT 1573 1607
 FT REPEAT 1608 1642
 FT REPEAT 1643 1677
 FT REPEAT 1678 1712
 FT REPEAT 1713 1747
 FT REPEAT 1748 1782
 FT REPEAT 1783 1817
 FT REPEAT 1818 1852
 FT REPEAT 1853 1887
 FT REPEAT 1888 1922
 FT REPEAT 1923 1957
 FT REPEAT 1958 1992
 FT REPEAT 1993 2027
 FT REPEAT 2028 2062
 FT REPEAT 2063 2097
 FT REPEAT 2098 2132
 FT REPEAT 2133 2167
 FT REPEAT 2168 2202
 FT REPEAT 2203 2237
 FT REPEAT 2238 2272
 FT REPEAT 2273 2307
 FT REPEAT 2308 2342
 FT REPEAT 2343 2377
 FT REPEAT 2378 2412
 FT REPEAT 2413 2447
 FT REPEAT 2448 2482
 FT REPEAT 2483 2517
 FT REPEAT 2518 2552
 FT REPEAT 2553 2587
 FT REPEAT 2588 2622
 FT REPEAT 2623 2657
 FT REPEAT 2658 2692
 FT REPEAT 2693 2727
 FT REPEAT 2728 2762
 FT REPEAT 2763 2797
 FT REPEAT 2798 2832
 FT REPEAT 2833 2867
 FT REPEAT 2868 2902
 FT REPEAT 2903 2937
 FT REPEAT 2938 2972
 FT REPEAT 2973 3007
 FT REPEAT 3008 3042
 FT REPEAT 3043 3077
 FT REPEAT 3078 3112
 FT REPEAT 3113 3147
 FT REPEAT 3148 3182
 FT REPEAT 3183 3217
 FT REPEAT 3218 3252
 FT REPEAT 3253 3287
 FT REPEAT 3288 3322
 FT REPEAT 3323 3357
 FT REPEAT 3358 3392
 FT REPEAT 3393 3427
 FT REPEAT 3428 3462
 FT REPEAT 3463 3497
 FT REPEAT 3498 3532
 FT REPEAT 3533 3567
 FT REPEAT 3568 3602
 FT REPEAT 3603 3637
 FT REPEAT 3638 3672
 FT REPEAT 3673 3707
 FT REPEAT 3708 3742
 FT REPEAT 3743 3777
 FT REPEAT 3778 3812
 FT REPEAT 3813 3847
 FT REPEAT 3848 3882
 FT REPEAT 3883 3917
 FT REPEAT 3918 3952
 FT REPEAT 3953 3987
 FT REPEAT 3988 4022
 FT REPEAT 4023 4057
 FT REPEAT 4058 4092
 FT REPEAT 4093 4127
 FT REPEAT 4128 4162
 FT REPEAT 4163 4197
 FT REPEAT 4198 4232
 FT REPEAT 4233 4267
 FT REPEAT 4268 4302
 FT REPEAT 4303 4337
 FT REPEAT 4338 4372
 FT REPEAT 4373 4407
 FT REPEAT 4408 4442
 FT REPEAT 4443 4477
 FT REPEAT 4478 4512
 FT REPEAT 4513 4547
 FT REPEAT 4548 4582
 FT REPEAT 4583 4617
 FT REPEAT 4618 4652
 FT REPEAT 4653 4687
 FT REPEAT 4688 4722
 FT REPEAT 4723 4757
 FT REPEAT 4758 4792
 FT REPEAT 4793 4827
 FT REPEAT 4828 4862
 FT REPEAT 4863 4897
 FT REPEAT 4898 4932
 FT REPEAT 4933 4967
 FT REPEAT 4968 5002
 FT REPEAT 5003 5037
 FT REPEAT 5038 5072
 FT REPEAT 5073 5107
 FT REPEAT 5108 5142
 FT REPEAT 5143 5177
 FT REPEAT 5178 5212
 FT REPEAT 5213 5247
 FT REPEAT 5248 5282
 FT REPEAT 5283 5317
 FT REPEAT 5318 5352
 FT REPEAT 5353 5387
 FT REPEAT 5388 5422
 FT REPEAT 5423 5457
 FT REPEAT 5458 5492
 FT REPEAT 5493 5527
 FT REPEAT 5528 5562
 FT REPEAT 5563 5597
 FT REPEAT 5598 5632
 FT REPEAT 5633 5667
 FT REPEAT 5668 5702
 FT REPEAT 5703 5737
 FT REPEAT 5738 5772
 FT REPEAT 5773 5807
 FT REPEAT 5808 5842
 FT REPEAT 5843 5877
 FT REPEAT 5878 5912
 FT REPEAT 5913 5947
 FT REPEAT 5948 5982
 FT REPEAT 5983 6017
 FT REPEAT 6018 6052
 FT REPEAT 6053 6087
 FT REPEAT 6088 6122
 FT REPEAT 6123 6157
 FT REPEAT 6158 6192
 FT REPEAT 6193 6227
 FT REPEAT 6228 6262
 FT REPEAT 6263 6297
 FT REPEAT 6298 6332
 FT REPEAT 6333 6367
 FT REPEAT 6368 6402
 FT REPEAT 6403 6437
 FT REPEAT 6438 6472
 FT REPEAT 6473 6507
 FT REPEAT 6508 6542
 FT REPEAT 6543 6577
 FT REPEAT 6578 6612
 FT REPEAT 6613 6647
 FT REPEAT 6648 6682
 FT REPEAT 6683 6717
 FT REPEAT 6718 6752
 FT REPEAT 6753 6787
 FT REPEAT 6788 6822
 FT REPEAT 6823 6857
 FT REPEAT 6858 6892
 FT REPEAT 6893 6927
 FT REPEAT 6928 6962
 FT REPEAT 6963 6997
 FT REPEAT 6998 7032
 FT REPEAT 7033 7067
 FT REPEAT 7068 7102
 FT REPEAT 7103 7137
 FT REPEAT 7138 7172
 FT REPEAT 7173 7207
 FT REPEAT 7208 7242
 FT REPEAT 7243 7277
 FT REPEAT 7278 7312
 FT REPEAT 7313 7347
 FT REPEAT 7348 7382
 FT REPEAT 7383 7417
 FT REPEAT 7418 7452
 FT REPEAT 7453 7487
 FT REPEAT 7488 7522
 FT REPEAT 7523 7557
 FT REPEAT 7558 7592
 FT REPEAT 7593 7627
 FT REPEAT 7628 7662
 FT REPEAT 7663 7697
 FT REPEAT 7698 7732
 FT REPEAT 7733 7767
 FT REPEAT 7768 7802
 FT REPEAT 7803 7837
 FT REPEAT 7838 7872
 FT REPEAT 7873 7907
 FT REPEAT 7908 7942
 FT REPEAT 7943 7977
 FT REPEAT 7978 8012
 FT REPEAT 8013 8047
 FT REPEAT 8048 8082
 FT REPEAT 8083 8117
 FT REPEAT 8118 8152
 FT REPEAT 8153 8187
 FT REPEAT 8188 8222
 FT REPEAT 8223 8257
 FT REPEAT 8258 8292
 FT REPEAT 8293 8327
 FT REPEAT 8328 8362
 FT REPEAT 8363 8397
 FT REPEAT 8398 8432
 FT REPEAT 8433 8467
 FT REPEAT 8468 8502
 FT REPEAT 8503 8537
 FT REPEAT 8538 8572
 FT REPEAT 8573 8607
 FT REPEAT 8608 8642
 FT REPEAT 8643 8677
 FT REPEAT 8678 8712
 FT REPEAT 8713 8747
 FT REPEAT 8748 8782
 FT REPEAT 8783 8817
 FT REPEAT 8818 8852
 FT REPEAT 8853 8887
 FT REPEAT 8888 8922
 FT REPEAT 8923 8957
 FT REPEAT 8958 8992
 FT REPEAT 8993 9027
 FT REPEAT 9028 9062
 FT REPEAT 9063 9097
 FT REPEAT 9098 9132
 FT REPEAT 9133 9167
 FT REPEAT 9168 9202
 FT REPEAT 9203 9237
 FT REPEAT 9238 9272
 FT REPEAT 9273 9307
 FT REPEAT 9308 9342
 FT REPEAT 9343 9377
 FT REPEAT 9378 9412
 FT REPEAT 9413 9447
 FT REPEAT 9448 9482
 FT REPEAT 9483 9517
 FT REPEAT 9518 9552
 FT REPEAT 9553 9587
 FT REPEAT 9588 9622
 FT REPEAT 9623 9657
 FT REPEAT 9658 9692
 FT REPEAT 9693 9727
 FT REPEAT 9728 9762
 FT REPEAT 9763 9797
 FT REPEAT 9798 9832
 FT REPEAT 9833 9867
 FT REPEAT 9868 9902
 FT REPEAT 9903 9937
 FT REPEAT 9938 9972
 FT REPEAT 9973 10007
 FT REPEAT 10008 10042
 FT REPEAT 10043 10077
 FT REPEAT 10078 10112
 FT REPEAT 10113 10147
 FT REPEAT 10148 10182
 FT REPEAT 10183 10217
 FT REPEAT 10218 10252
 FT REPEAT 10253 10287
 FT REPEAT 10288 10322
 FT REPEAT 10323 10357
 FT REPEAT 10358 10392
 FT REPEAT 10393 10427
 FT REPEAT 10428 10462
 FT REPEAT 10463 10497
 FT REPEAT 10498 10532
 FT REPEAT 10533 10567
 FT REPEAT 10568 10602
 FT REPEAT 10603 10637
 FT REPEAT 10638 10672
 FT REPEAT 10673 10707
 FT REPEAT 10708 10742
 FT REPEAT 10743 10777
 FT REPEAT 10778 10812
 FT REPEAT 10813 10847
 FT REPEAT 10848 10882
 FT REPEAT 10883 10917
 FT REPEAT 10918 10952
 FT REPEAT 10953 10987
 FT REPEAT 10988 11022
 FT REPEAT 11023 11057
 FT REPEAT 11058 11092
 FT REPEAT 11093 11127
 FT REPEAT 11128 11162
 FT REPEAT 11163 11197
 FT REPEAT 11198 11232
 FT REPEAT 11233 11267
 FT REPEAT 11268 11302
 FT REPEAT 11303 11337
 FT REPEAT 11338 11372
 FT REPEAT 11373 11407
 FT REPEAT 11408 11442
 FT REPEAT 11443 11477
 FT REPEAT 11478 11512
 FT REPEAT 11513 11547
 FT REPEAT 11548 11582
 FT REPEAT 11583 11617
 FT REPEAT 11618 11652
 FT REPEAT 11653 11687
 FT REPEAT 11688 11722
 FT REPEAT 11723 11757
 FT REPEAT 11758 11792
 FT REPEAT 11793 11827
 FT REPEAT 11828 11862
 FT REPEAT 11863 11897
 FT REPEAT 11898 11932
 FT REPEAT 11933 11967
 FT REPEAT 11968 12002
 FT REPEAT 12003 12037
 FT REPEAT 12038 12072
 FT REPEAT 12073 12107
 FT REPEAT 12108 12142
 FT REPEAT 12143 12177
 FT REPEAT 12178 12212
 FT REPEAT 12213 12247
 FT REPEAT 12248 12282
 FT REPEAT 12283 12317
 FT REPEAT 12318 12352
 FT REPEAT 12353 12387
 FT REPEAT 12388 12422
 FT REPEAT 12423 12457
 FT REPEAT 12458 12492
 FT REPEAT 12493 12527
 FT REPEAT 12528 12562
 FT REPEAT 12563 12597
 FT REPEAT 12598 12632
 FT REPEAT 12633 12667
 FT REPEAT 12668 12702
 FT REPEAT 12703 12737
 FT REPEAT 12738 12772
 FT REPEAT 12773 12807
 FT REPEAT 12808 12842
 FT REPEAT 12843 12877
 FT REPEAT 12878 12912
 FT REPEAT 12913 12947
 FT REPEAT 12948 12982
 FT REPEAT 12983 13017
 FT REPEAT 13018 13052
 FT REPEAT 13053 13087
 FT REPEAT 13088 13122
 FT REPEAT 13123 13157
 FT REPEAT 13158 13192
 FT REPEAT 13193 13227
 FT REPEAT 13228 13262
 FT REPEAT 13263 13297
 FT REPEAT 13298 13332
 FT REPEAT 13333 13367
 FT REPEAT 13368 13402
 FT REPEAT 13403 13437
 FT REPEAT 13438 13472
 FT REPEAT 13473 13507
 FT REPEAT 13508 13542
 FT REPEAT 13543 13577
 FT REPEAT 13578 13612
 FT REPEAT 13613 13647
 FT REPEAT 13648 13682
 FT REPEAT 13683 13717
 FT REPEAT 13718 13752
 FT REPEAT 13753 13787
 FT REPEAT 13788 13822
 FT REPEAT 13823 13857
 FT REPEAT 13858 13892
 FT REPEAT 13893 13927
 FT REPEAT 13928 13962
 FT REPEAT 13963 13997
 FT REPEAT 13998 14032
 FT REPEAT 14033 14067
 FT REPEAT 14068 14102
 FT REPEAT 14103 14137
 FT REPEAT 14138 14172
 FT REPEAT 14173 14207
 FT REPEAT 14208 14242
 FT REPEAT 14243 14277
 FT REPEAT 14278 14312
 FT REPEAT 14313 14347
 FT REPEAT 14348 14382
 FT REPEAT 14383 14417
 FT REPEAT 14418 14452
 FT REPEAT 14453 14487
 FT REPEAT 14488 14522
 FT REPEAT 14523 14557
 FT REPEAT 14558 14592
 FT REPEAT 14593 14627
 FT REPEAT 14628 14662
 FT REPEAT 14663 14697
 FT REPEAT 14698 14732
 FT REPEAT 14733 14767
 FT REPEAT 14768 14802
 FT REPEAT 14803 14837
 FT REPEAT 14838 14872
 FT REPEAT 14873 14907
 FT REPEAT 14908 14942
 FT REPEAT 14943 14977
 FT REPEAT 14978 15012
 FT REPEAT 15013 15047
 FT REPEAT 15048 15082
 FT REPEAT 15083 15117
 FT REPEAT 15118 15152
 FT REPEAT 15153 15187
 FT REPEAT 15188 15222
 FT REPEAT 15223 15257
 FT REPEAT 15258 15292
 FT REPEAT 15293 15327
 FT REPEAT 15328 15362
 FT REPEAT 15363 15397
 FT REPEAT 15398 15432
 FT REPEAT 15433 15467
 FT REPEAT 15468 15502
 FT REPEAT 15503 15537
 FT REPEAT 15538 15572
 FT REPEAT 15573 15607
 FT REPEAT 15608 15642
 FT REPEAT 15643 15677
 FT REPEAT 15678 15712
 FT REPEAT 15713 15747
 FT REPEAT 15748 15782
 FT REPEAT 15783 15817
 FT REPEAT 15818 15852
 FT REPEAT 15853 15887
 FT REPEAT 15888 15922
 FT REPEAT 15923 15957
 FT REPEAT 15958 15992
 FT REPEAT 15993 16027
 FT REPEAT 16028 16062
 FT REPEAT 16063 16097
 FT REPEAT 16098 16132
 FT REPEAT 16133 16167
 FT REPEAT 16168 16202
 FT REPEAT 16203 16237
 FT REPEAT 16238 16272
 FT REPEAT 16273 16307
 FT REPEAT 16308 16342
 FT REPEAT 16343 16377
 FT REPEAT 16378 16412
 FT REPEAT 16413 16447
 FT REPEAT 16448 16482
 FT REPEAT 16483 16517
 FT REPEAT 16518 16552
 FT REPEAT 16553 16587
 FT REPEAT 16588 16622
 FT REPEAT 16623 16657
 FT REPEAT 16658 16692
 FT REPEAT 16693 16727
 FT REPEAT 16728 16762
 FT REPEAT 16763 16797
 FT REPEAT 16798 16832
 FT REPEAT 16833 16867
 FT REPEAT 16868 16902
 FT REPEAT 16903 16937
 FT REPEAT 16938 16972
 FT REPEAT 16973 17007
 FT REPEAT 17008 17042
 FT REPEAT 17043 17077
 FT REPEAT 17078 17112
 FT REPEAT 17113 17147
 FT REPEAT 17148 17182
 FT REPEAT 17183 17217
 FT REPEAT 17218 17252
 FT REPEAT 17253 17287
 FT REPEAT 17288 17322
 FT REPEAT 17323 17357
 FT REPEAT 17358 17392
 FT REPEAT 17393 17427
 FT REPEAT 17428 17462
 FT REPEAT 17463 17497
 FT REPEAT 17498 17532
 FT REPEAT 17533 17567
 FT REPEAT 17568 17602
 FT REPEAT 17603 17637
 FT REPEAT 17638 17672
 FT REPEAT 17673 17707
 FT REPEAT 17708 17742
 FT REPEAT 17743 17777
 FT REPEAT 17778 17812
 FT REPEAT 17813 17847
 FT REPEAT 17848 17882
 FT REPEAT 17883 17917
 FT REPEAT 17918 17952
 FT REPEAT 17953 17987
 FT REPEAT 17988 18022
 FT REPEAT 18023 18057
 FT REPEAT 18058 18092
 FT REPEAT 18093 18127
 FT REPEAT 18128 18162
 FT REPEAT 18163 18197
 FT REPEAT 18198 18232
 FT REPEAT 18233 18267
 FT REPEAT 18268 18302
 FT REPEAT 18303 18337
 FT REPEAT 18338 18372
 FT REPEAT 18373 18407
 FT REPEAT 18408 18442
 FT REPEAT 18443 18477
 FT REPEAT 18478 18512
 FT REPEAT 18513 18547
 FT REPEAT 18548 18582
 FT REPEAT 18583 18617
 FT REPEAT 18618 18652
 FT REPEAT 18653 18687
 FT REPEAT 18688 18722
 FT REPEAT 18723 18757
 FT REPEAT 18758 18792
 FT REPEAT 18793 18827
 FT REPEAT 18828 18862
 FT REPEAT 18863 18897
 FT REPEAT 18898 18932
 FT REPEAT 18933 18967
 FT REPEAT 18968 19002
 FT REPEAT 19003 19037
 FT REPEAT 19038 19072
 FT REPEAT 19073 19107
 FT REPEAT 19108 19142
 FT REPEAT 19143 19177
 FT REPEAT 19178 19212
 FT REPEAT 19213 19247
 FT REPEAT 19248 19282
 FT REPEAT 19283 19317
 FT REPEAT 19318 19352
 FT REPEAT 19353 19387
 FT REPEAT 19388 19422
 FT REPEAT 19423 19457
 FT REPEAT 19458 19492
 FT REPEAT 19493 19527
 FT REPEAT 19528 19562
 FT REPEAT 19563 19597
 FT REPEAT 19598 19632
 FT REPEAT 19633 19667
 FT REPEAT 19668 19702
 FT REPEAT 19703 19737
 FT REPEAT 19738 19772
 FT REPEAT 19773 19807
 FT REPEAT 19808 19842
 FT REPEAT 19843 19877
 FT REPEAT 19878 19912
 FT REPEAT 19913 19947
 FT REPEAT 19948 19982
 FT REPEAT 19983 20017
 FT REPEAT

```

FT DISULFID 127 143 BY SIMILARITY.
FT DISULFID 145 154 BY SIMILARITY.
FT DISULFID 160 171 BY SIMILARITY.
FT DISULFID 165 180 BY SIMILARITY.
FT DISULFID 182 191 BY SIMILARITY.
FT DISULFID 198 211 BY SIMILARITY.
FT DISULFID 205 220 BY SIMILARITY.
FT DISULFID 222 231 BY SIMILARITY.
FT DISULFID 238 249 BY SIMILARITY.
FT DISULFID 243 262 BY SIMILARITY.
FT DISULFID 264 273 BY SIMILARITY.
FT DISULFID 280 291 BY SIMILARITY.
FT DISULFID 285 300 BY SIMILARITY.
FT DISULFID 302 311 BY SIMILARITY.
FT DISULFID 318 332 BY SIMILARITY.
FT DISULFID 326 341 BY SIMILARITY.
FT DISULFID 343 352 BY SIMILARITY.

Query Match 18.88; Score 675.5; DB 1; Length 2003;
Best Local Similarity 26.38; Pred. No. 6.3e-36;
Matches 211; Conservative 45; Mismatches 210; Indels 337; Gaps 46;

QY 94 CCGFYESEKCVPHCADKC-----VHCRIAPNT-----CQCEBGMGTNCSSACDGDH 143
    ||| :|| :|
Db 105 CLPGR--TGERCOAKLEDPCEPSPFSCSKRGCHIQASGRPCSCMGMTGEOQLR----- 157
    ||| :|| :|

QY 144 WGPCTRSCCKNGALCNPITG--ACHCAAFGRNRC--DRCEQS---TYGNDCHQ- 193
    ||| :|| :|
Db 158 --DICSAN-PCVNGVCLATTPOIQCPCPEFEGHACERDVNECFODPGCPKGTSCNHT 214
    ||| :|| :|

QY 194 -----RCQ-----CQNGATC-----DHYTECRCPPGYTGARCE 222
    ||| :|| :|
Db 215 LGSFQCLCPVQEGPRCLRAGPCPRRCNSNGTCQLMPEDSFTHLCLCPFGIPGCE 274
    ||| :|| :|

QY 223 -----DLCPRCKHG-----PCEQRCP--CQNGVCYCH 248
    ||| :|| :|
Db 275 VNPNCVSHOCNGSTGCGDGLDYTCLEPMTGMDCESEVDECTGPRHCRNGATCON 334
    ||| :|| :|

QY 249 VTG--ECCSPGMMGTVCQGP-----PCEGRFRNCSE- 280
    ||| :|| :|
Db 335 SAGSFHCVCVSGMGSTCEENLDCIATACSTCIDRNGSFCLCPGGTGLLCHLED 394
    ||| :|| :|

QY 281 -C---QCHNGGTC--DATGQ--CHCSPGYTERCQ---DECPYGTGVLCATRCQCVNG 329
    ||| :|| :|
Db 395 MCLSPCHGDAQCSSTNPLTGSTLCLCQPGSGPTCHQDDBCLMAQGS---PSRPHG 449
    ||| :|| :|

QY 330 GKCVHSGA--CLCEAFAGEBRCAR-----LCPEGLYGIK 363
    ||| :|| :|
Db 450 GSCILNTGFSNCLCPPGYTSRCEADHNECLSPCHGPGSTCLDLATFHCLCPGLBGL 509
    ||| :|| :|

QY 364 CD--KRC--PCHLEWTHSCHPMMSG--ECACKPGMSGLYCNE----- 398
    ||| :|| :|
Db 510 CEVETNCSAARPC--LNHACHDHLNFCQICLPFGSGTCEEDIDECRSSPCANGGCOQ 567
    ||| :|| :|

QY 399 -----TCSPGFYGACQ--QIC--SCONGADCDVYTGK--CTCAFGRKIDICSTP 442
    ||| :|| :|
Db 568 DQGAFFHCKCLPFGPRCQTEVDECLSDPCPGASGLDLPAGFELCLPBGFTGOLCEVP 627
    ||| :|| :|

QY 443 -----CPGTYTIT--NCSRCG--CKNDAYCSPVDSGCTC 473
    ||| :|| :|
Db 628 LCAFNLCOPKQIKKDKRANCLCPDPSGCAPEDEKCTHNGHQR-----SSCYVC 679
    ||| :|| :|

QY 474 KAGMHGVDG-----STNCPGWTGFGON--LNCQ---CL 502
    ||| :|| :|
Db 680 DVGWTPRECAELGGLISAPCANGGTCYPOPSGYNCTCFPGYPTCSSEMTATCHSGPCL 739
    ||| :|| :|

QY 503 NGACACTLDS--TCTCAPGRGECF-----LPC-----ODGTYGLNCA----- 539
    ||| :|| :|
Db 740 NGSSCNPPGPGYCTCPPSHTGPOCOTSDYDVCASAPCFNGTCVNRPGFTSCCLAMGFOG 799
    ||| :|| :|

QY 540 ERD-----CSHADGCH--PTTGICRCLCPMGSGVHCDS----- 570
    ||| :|| :|
Db 800 PRCEGLRPSCADSPCRNRATCODSPFGPRCLCPGTGYTGSGCQTLMDLCAQKPCPRNSHC 859
    ||| :|| :|

```

```

QY 571 -----VCAEGRNGPNCSLP 584
    :| :| :| :|
Db 860 IQTGPSFHCICLQGWGTGFLCNP 882

RESULT 13
NTCL_RAT STANDARD; PRT; 2531 AA.
AC Q07008;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor (Notch 1).
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weimaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian development.";
RL Development 113:199-205(1991).
RN [2]
RP REVISIONS TO 1652-1653.
RA Weimaster G.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP FUNCTION.
RX MEDLINE=21094508; PubMed=11182080;
RA Tanigaki K., Nogaki F., Takahashi J., Tashiro K., Kuwoka H.,
RT Honjo T.;
RL "Notch1 and Notch3 instructively restrict bFGF-responsive multipotent neural progenitor cells to an astroglial fate.";
RN [4]
RP TISSUE SPECIFICITY.
RX MEDLINE=93202015; PubMed=1295745;
RA Weimaster G., Roberts V.J., Lemke G.;
RT "Notch2: a second mammalian Notch gene.";
RL Development 116:931-941(1992).
RN [5]
RP TISSUE SPECIFICITY.
RX MEDLINE=21331789; PubMed=11438922;
RA Irvin D.K., Zurcher S.D., Nguyen T., Weimaster G., Kornblum H.I.;
RT "Expression patterns of Notch1, Notch2, and Notch3 suggest multiple functional roles for the Notch-DSL signaling system during brain development.";
RL J. Comp. Neurol. 436:167-181(2001).
RN [6]
RP FUNCTION. Functions as a receptor for membrane-bound ligands
CC jagged1, jagged2 and Deltal to regulate cell-fate determination.
CC upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). Acts instructively to control
CC the cell fate determination of CNS multipotent progenitor cells,
CC resulting in astroglial induction and neuron/oligodendrocyte
CC suppression.
CC [7]
CC SUBUNIT: Heterodimer of a C-terminal fragment N(Tm) and a N-
CC terminal fragment N(FC) which are probably linked by disulfide
CC bonds (By similarity).
CC [8]
CC SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus (By
CC similarity).
CC [9]
CC TISSUE SPECIFICITY: Expressed in the brain, kidney and spleen.
CC Expressed in postnatal central nervous system (CNS) germinal zones
CC and in early postnatal life, within numerous cells throughout the
CC CNS. Found in both subventricular and ventricular germinal zones.
CC

```



```

Db 647 SNPCDS---GTCLDKIDGYECA-----CEPGY--TGSNCNVN 678
OY 106 VPHCA-----DKCVHGRG---TAPNT 123
Db 679 IDCAASPCINGSTCEDJAGFTCRCPREGYHDPCLSEVNECHSNPCIRHACADGLNGYX 738
OY 124 COCEPGMGTCN---SSACDGDHMGPHCTSRCCCKNGALCNPTG---ACHCAAGFRGWRG 178
Db 739 CDCAPMGSTGCDINNNECESN-----PCVNGGTCKDMTSGYVTCRGFGSPGNC 788
OY 179 EDCEGEGTGNDCHORCCQONGATC-DHYTG-ECRCRPGTGAFCEDLCPRGKHGPOCEQ 236
Db 789 Q-----TINECASN-PCLNQGTCTDVAGYKNCNCPPLPTGATCEVYLA-----C-A 834
OY 237 RCPGONGVC---HHVTGECSCPSGMMGTVC-----GQCPPEGR-----272
Db 835 TSPCKNSGCKESEDYESFCVPTMGQGTCEIDINECVKSPCRGASQONTNGSYRL 894
OY 273 -----FGKNCN---OECQ---CHNGTCDAAATGQ---CHCSPGYTGRCODE-----310
Db 895 COAGYTRNCESDIDCCRPMPCHNGSGCTDGVNAAFCDCLPGFQAFCEEDINECATNPG 954
OY 311 -----CPVTFYGLCAET-----GQVNGSKCYHVS---ACICEAG 344
Db 955 QNCACTDCVDSYTCPTGPFNGIHCBENNTPDCTESSCFNGGTC---VDGINSFTCLCPPG 1012
OY 345 FAGERCEARLCPRGLYGI-KCDARCPCHLENTSCHPMG--ECACKPGMSGLYCNE---398
Db 1013 FTGSYCO-----YDVNCDNR-PLHNGT---CODSYGTGCTCPGTGTGLNQNLVR 1061
OY 399 --TCSGFGFGEACQOI-----GSCQN---GACDDSYTGKCTCAPRGKIDCSFPCLGY 448
Db 1062 WCDASACKNGKGMQNTNTYHCECRSGMTGFNDVLSVSCVAAORGRGIDVTLLQHGGL 1121
OY 449 GIN-----CSSRCG-----CKNDVAVSPVGG--SCTKAGMHGVDCS 483
Db 1122 CUDEBKHYCHCQAGTGTSGCEDEBECSPNQCQNGATCTDYGSGCKCAVATHSNCS 1181
OY 484 -----IRCPSGTWGFGCNLT---C-----OCLNG 504
Db 1182 EEINELSLPONGGTCTIDLTNTYKCSPRGTGVCHEINVDCHPLDPAASRSPCFNN 1241
OY 505 GACNTIDG--TCTCARGMREKE-----LPCOD-GTYGLNCAERCCSHADGCHPT 533
Db 1242 GTCVDGVGTCTCPGPGFGEERCEGDVNECLSNPCDRGTQ--NCQVRN-----1289
OY 554 TGHCRCLPGMSGVHCDVCAEGRMGPC 581
Db 1290 DFHCERAGHTGRCESV--INGCRKPC 1316

```

```

CC fate decisions.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -1- SIMILARITY: CONTAINS 17 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DSL DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 VWF C DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF229451; AL08216.1; -.
CC ZFIN: ZDB-GENE-011128-4; JAG3.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR001774; DSL.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR00742; EGF_2.
CC InterPro: IPR01881; EGF_Ca.
CC InterPro: IPR01007; VWF_C.
CC Pfam: PF01414; DSL; 1.
CC Pfam: PF00008; EGF; 18.
CC Pfam: PF00093; VWC; 1.
CC PROSITE: PS00010; ASX_HYDROXYL; 10.
CC PROSITE: PS00022; EGF_1; 16.
CC PROSITE: PS01186; EGF_2; 12.
CC PROSITE: PS01187; EGF_Ca; 8.
CC PROSITE: PS01208; VWF_C; FALSE_NEG.
CC KEGG: Calcium-binding; EGF-like domain; Glycoprotein; Developmental protein;
CC Repeat; Transmembrane; Signal.
CC KW CHAIN 1 27 1213
CC FT SIGNAL 27 1064
CC FT DOMAIN 1065 1087
CC FT TRANSMEM 1088 1213
CC FT DOMAIN 164 226
CC FT DOMAIN 227 260
CC FT DOMAIN 258 291
CC FT DOMAIN 293 331
CC FT DOMAIN 332 369
CC FT DOMAIN 371 407
CC FT DOMAIN 409 445
CC FT DOMAIN 447 482
CC FT DOMAIN 484 520
CC FT DOMAIN 522 558
CC FT DOMAIN 559 624
CC FT DOMAIN 626 662
CC FT DOMAIN 664 700
CC FT DOMAIN 702 738
CC FT DOMAIN 746 777
CC FT DOMAIN 779 815
CC FT DOMAIN 817 853
CC FT DOMAIN 860 914
CC FT DOMAIN 918 956
CC FT DSLFID 231 242
CC FT DSLFID 235 248
CC FT DSLFID 250 259
CC FT DSLFID 262 273
CC FT DSLFID 268 279
CC FT DSLFID 281 290
CC FT DSLFID 297 309
CC FT DSLFID 303 319
CC FT DSLFID 321 330
CC FT DSLFID 337 348
CC FT DSLFID 342 357
CC FT DSLFID 359 368
CC FT DSLFID 375 386
CC FT DSLFID 380 395
CC FT DSLFID 397 406
CC FT DSLFID 413 424
CC FT DSLFID 418 433

```



```

FT DISULFID 435 444 BY SIMILARITY.
FT DISULFID 451 461 BY SIMILARITY.
FT DISULFID 455 470 BY SIMILARITY.
FT DISULFID 472 481 BY SIMILARITY.
FT DISULFID 488 499 BY SIMILARITY.
FT DISULFID 493 508 BY SIMILARITY.
FT DISULFID 510 519 BY SIMILARITY.
FT DISULFID 526 537 BY SIMILARITY.
FT DISULFID 531 546 BY SIMILARITY.
FT DISULFID 548 557 BY SIMILARITY.
FT DISULFID 630 641 BY SIMILARITY.
FT DISULFID 635 650 BY SIMILARITY.
FT DISULFID 652 661 BY SIMILARITY.
FT DISULFID 668 679 BY SIMILARITY.
FT DISULFID 673 688 BY SIMILARITY.
FT DISULFID 690 699 BY SIMILARITY.
FT DISULFID 706 717 BY SIMILARITY.
FT DISULFID 711 726 BY SIMILARITY.
FT DISULFID 728 737 BY SIMILARITY.
FT DISULFID 783 794 BY SIMILARITY.
FT DISULFID 788 803 BY SIMILARITY.
FT DISULFID 805 814 BY SIMILARITY.
FT DISULFID 821 832 BY SIMILARITY.
FT DISULFID 826 841 BY SIMILARITY.
FT DISULFID 843 852 BY SIMILARITY.
FT DOMAIN 938 941 POLY-PRO.
FT CARBOHD 139 139 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 214 214 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 742 742 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 957 957 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 988 988 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHD 1042 1042 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1213 AA; 133365 MW; 5C5F16A7E20D9534 CRC64;

```

Query Match 18.5%; Score 667.5; DB 1; Length 1213;
 Best Local Similarity 27.7%; Pred. No. 1.4e-35;
 Matches 190; Conservative 64; Mismatches 176; Indels 257; Gaps 45;

```

OY 83 GKTYRRRSQCCPFYEGEGCVHCHADKCYHRCIAPIANTQCCPBGNGTSSACDGD 142
DB 247 GKCYLGMWGPYC-----DKCIPH--PGVHGTCVEFMQCLDPTNMWG-----OLCDXD 293
OY 143 HMGPHCTSRCCCKNGALCN---PITGACHCAAGFRGMRCE-----DRCEGG 185
DB 294 L--NVCGRHQPCLNGTSGNTGRDYQSCEDGYGVNCEBHAHCLSPCANGTCKET 351
OY 186 TYGNDCH-----QRC---OCQNGATC-DHYTG-ECRCPPGYTGAFCEDLCP 226
DB 352 SQGYEHCALIGWSTSCETINVDCTPNQCKHGGTCQDLVNGFKACAPPHMTGKTCQ---- 407
OY 227 PEKHRCQEQRCQNGCVGCHVTGE--CSCPSGMMGYTCGPPCBGRGRKNC---QEC 281
DB 408 --IDANECDK-PCVYNAKSNHNLGAYFCECLPGWS-----GQNCDDININDC 451
OY 282 --QCHNGGTC-DAATG-OCHCSPGYTGERCQ---DECPVGYGVLAETCCQVNGKCY- 333
DB 452 KGQCLNGGCKDLVNGYRKLCPRGYTGERCEKDVEDC-----ASSPLNGRGROD 501
OY 334 HVSQ-ACLCBAGFAGEKCEARL-----CPBG--LYGIKCDKRCPC-----HLE 373
DB 502 EYNGEQCLCPAGFSGQLDIDYCKPMPQNGACQCFNLASDYPFCPCPDYEGKNCSHLK 561
OY 374 N---THSCHPMGSG-----EACRKGMSGL 394
DB 562 DHCRITSCQVIDSTVAANASTPEGVYRISSNVCGPHGRCSRQAGQFTCCQGGFRGT 621
OY 395 YCNETCSPEFYGEACQOICSCONGADC-DSV-TGKCTCAPGFKGIDSTPPIAGTYGINC 452
DB 622 YCHEININD-----CES-NPCRNNGTCLDKVAVNYCCICADGEGVHCE-----INDD 667
OY 453 SSRGCKNDAYCSPV--DGSCTCKAGMHGVDCSIF--CPSGTWFGFNGNLTCQCLNGAGC 507

```

```

DB 668 CSLNPLNKAGACQDLYNDFYCECRNGWKGTCHSDSODEAT-----CNNGGTC 717
OY 508 NTLDDGT--CTCAPGRNGECE-----LP-----CODGTGLMCA 539
DB 718 HDGDTFKRCRSPNEGATCNIAKNSCLLPNCENGGTCVWNGDSFNVCYCKEGMGSTCT 777
OY 540 ERC-DCSHADGCHP--TTG-----HCRCLPMSGVHC-----568
DB 778 ENTNDQNP-----HPCYNSGTCTVDGENWYRCBAPGAPGDCININTEGSSPCARGSTCV 833
OY 569 DSV-----CAEGRNMPNC-----SLPC 585
DB 834 DEINGYRCICPPGRIGIPDQCEVYGRPC 860

RESULT 15
NTC2_HUMAN
ID NTC2_HUMAN STANDARD; PRG: 2471 AA.
AC 004721; Q9H240; Q99734;
DT 15-JUN-2002 (Rel. 41, Created)
DI 15-JUN-2002 (Rel. 41, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (hN2).
GN NOTCH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain;
RA Blumweller C.M., Mann R.S.;
RT "Complete human notch 2 (hN2) cDNA sequence.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-Breast tumor;
RA Correa R.G., Camargo A.A., Moreira E.S., Simpson A.J.;
RT "Human Notch2, a novel member of cell-fate determining NOTCH
family.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-T-cell;
RA Lemasson I., Devaux C., Mesnard J.M.;
RT "Partial sequence of EGF-like repeat domain of human Notch2 mRNA.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1810-2447 FROM N.A.
RC TISSUE-Brain;
RA MEDLINE=93265135; PubMed=1303260;
RA Stifani S., Blumweller C.M., Redhead N.J., Hall R.E.,
RA Artavanis-Tsakonas S.;
RT "Human homologs of a Drosophila enhancer of split gene product define
a novel family of nuclear proteins.";
RN [5]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=97386453; PubMed=9244302;
RA Blumweller C.M., Qi H., Zagouras P., Artavanis-Tsakonas S.;
RT "Intracellular cleavage of Notch leads to a heterodimeric receptor on
the plasma membrane.";
RL Cell 90:281-291(1997).
RN [6]
RP IDENTIFICATION OF LIGANDS.
RX MEDLINE=99180765; PubMed=10079256;
RA Gray G.E., Mann R.S., Mitsiadis E., Henrique D., Carcangiu M.-D.,
RA Barks A., Leiman J., Ward D., Ish-Horowitz D., Artavanis-Tsakonas S.;
RT "Human ligands of the Notch receptor.";
RL Am. J. Pathol. 154:785-794(1999).
CC -I- FUNCTION: Functions as a receptor for membrane-bound ligands
jagged1, jagged2 and Delta1 to regulate cell-fate determination.
upon ligand activation through the released notch intracellular

```


THIS PAGE BLANK (USPTO)